# MODELO DE REGRESSÃO ELÍPTICO COM ERROS NAS VARIÁVEIS

**Filidor Edilfonso Vilca Labra**

TESE APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTENÇÃO DO GRAU
DE
DOUTOR EM ESTATÍSTICA

Área de Concentração: **Estatística**
Orientador: **Prof. Dr. Heleno Bolfarine**
Co-Orientador: **Prof. Dr. Reinaldo Boris Arellano Valle**

– São Paulo, junho de 1996 –

MODELO DE REGRESSÃO ELÍPTICO
COM ERROS NAS VARIÁVEIS

Filidor Edilfonso Vilca Labra

Este exemplar corresponde à redação final da tese apresentada por Filidor Edilfonso Vilca Labra, devidamente corrigida e aprovada pela Comissão Julgadora.

São Paulo, 06 de setembro de 1996

**Comissão Julgadora**

- Prof. Dr. Heleno Bolfarine - IME-USP
- Prof. Dr. José Galvão Leite - IME-USP
- Prof. Dr. Mauro Sérgio de Freitas Marques - UNICAMP
- Prof. Dr. Gauss Moutinho Cordeiro - UFPE
- Profa. Dra. Chang Chung Yu Dorea - UnB

# AGRADECIMENTOS

Ao Prof. Dr. Heleno Bolfarine, pela orientação amiga, pelo incentivo constante e sobretudo pela confiança em mim depositada durante o desenvolvimento deste trabalho.

Ao Prof. Dr. Reinaldo Boris Arellano Valle pela co-orientação, cujas sugestões foram fundamentais na realização deste trabalho.

Ao Prof. Dr. José Galvão Leite, pelo incentivo e boa disposição para discutir vários aspectos deste trabalho.

A meus amigos e colegas: Antonio, Oscar, Miguel, Ivan, Manuel, Patrícia, Wânia, Manoel e Cláudia, pela amizade, companheirismo e boa disposição que sempre tiveram comigo nos momentos difíceis.

A meus tios, Plácido Roque e Felisa Labra pelo apoio nos momentos difíceis de minha vida. A meus irmãos Carlos e Haydee, em especial a meu irmão Pedro pelo incentivo e constante preocupação.

Agradeço de forma especial à Profa. Iracema, por seu apoio e preocupação.

Quero agradecer ao Walter, por sua paciência e notável eficiência na digitação desta tese.

A CAPES, pelo apoio financeiro, que permitiu a realização de meus estudos neste país.

Agradeço finalmente a Deus, por ter possibilitado todos os agradecimentos acima.

# RESUMO

Neste trabalho estudamos inferência em modelos com erros nas variáveis considerando que as variáveis do modelo são distribuídas de acordo com uma distribuição da família das distribuições elípticas. O estudo é desenvolvido para os modelos ultraestrutural, estrutural e funcional. Como conseqüência, os resultados obtidos na literatura sob normalidade são generalizados.

No modelo estrutural, discutimos o comportamento assintótico do estimador de máxima verossimilhança e a eficiência assintótica relativa deste em relação a outros estimadores. Além disso, no modelo consistindo de duas ou mais populações estruturais, abordamos o problema de inferência para os coeficientes angulares do modelo.

No modelo funcional, estudamos o comportamento assintótico do estimador de máxima verossimilhança com especial atenção a um modelo freqüentemente usado em estudos de calibração comparativa.

# ABSTRACT

In this thesis, we consider inference problems in measurement error models assuming that the variables involved follow an elliptical distribution. The investigation concerns ultraestructural, structural and functional models. As a byproduct of the study, some results obtained under normality are generalized. In the structural model, the asymptotic behaviour of the maximum likelihood estimator of the slope parameter is considered and the asymptotic relative efficiencies with respect to other competitors are studied. This study is extended to the case where two or more structural populations are under investigation.

In the functional model, the asymptotic behaviour of the maximum likelihood estimator is investigated, with special attention to the comparative calibration model, typically considered when comparing several measuring instruments.

# Conteúdo

# Capítulo 1

# Introdução

## 1.1 Formulação do problema

A análise clássica multivariada de dados contínuos está centrada basicamente na distribuição normal multivariada. Este fato é habitualmente justificado via o Teorema Central do Limite, razão pela qual a maior parte da teoria da inferência estatística está desenvolvida em torno deste modelo como, por exemplo, os modelos lineares. Contudo, em muitas situações o modelo sob normalidade é impróprio como, por exemplo, quando os dados provêm de uma distribuição com caudas mais pesadas que a distribuição normal. De fato, a vulnerabilidade da inferência baseada no modelo normal com respeito a observações aberrantes ("outliers") tem sido motivo de vários estudos de robustez. Modelos alternativos ao modelo normal que preservam sua estrutura simétrica e que permitem reduzir a influência dos "outliers", têm sido sugeridos por diferentes autores. Lange, Little e Taylor (1989), por exemplo, propõem o modelo $t$-Student como uma extensão paramétrica robusta do modelo normal, e ilustram o desempenho do novo modelo numa grande variabilidade de aplicações. Os trabalhos recentes de robustez, mostram que a inferência estatística clássica, desenvolvida para um modelo normal, é basicamente correta, não somente para o modelo normal em questão, mas também para modelos que provêm de distribuições elípticas, desde que um certos ajustes sejam feitos, quando é necessário (Muirhead e Waternaux, 1980, Browne 1982 e 1984, Bentler, 1983, Shapiro e Browne, 1987, e Kano, Berkane e Bentler, 1993). Contudo, Tyler (1983) nota que sendo a distribuição da população elíptica, a inferência baseada na teoria estatística normal é significativamente ineficiente quando a curtose da distribuição elíptica assume valores bem mais altos. Isto sugere a necessidade de desenvolver inferências numa determinada classe de modelos elípticos que possam cobrir distribuições com curtose alta, ao invés de realizar ajustes na teoria de inferência normal. Vários autores têm sugerido a utilização de distribuições com caudas mais pesadas, entre os quais, destacamos Lange et al. (1989) que recomendam a distribuição $t$-Student, e Little (1988) e Yamaguchi (1990) que usam a distribuição normal contaminada. Ambos os modelos incorporam parâmetros adicionais que permitem ajustar a curtose da distribuição dos dados. No caso da distribuição $t$-Student, a curtose é modelada através dos graus de liberdade e portanto inclui em particular as distribuições Cauchy e normal.

O estudo teórico e aplicado das distribuições elípticas, como é o caso da distribuição $t$-Student, começou a desenvolver-se com muito interesse a partir do trabalho devido a Kelker (1970). Desde então, estas distribuições têm sido objeto de diversos estudos, tanto no aspecto teórico quanto em diversas áreas de aplicação. Alguns resultados neste sentido são encontrados, por exemplo, em Devlin, Gnadesikan e Kettenring (1976), Kariya e Eaton (1977), Muirhead e Waternaux (1980) e Muirhead (1980). Destaca-se o excelente trabalho de Cambanis, Huang e Simons (1981) que proporciona um rigoroso e sistemático tratamento probabilístico das distribuições elípticas e também o trabalho de Anderson, Fang e Hsu (1986) deriva resultados gerais sobre os EMV no modelo eliptico dependente. A maioria destes trabalhos pode ser encontrada de forma unificada em Fang, Kotz e Ng (1990). A classe das distribuições elípticas proporciona uma grande variedade de modelos estatísticos alternativos ao modelo normal e em tais modelos, diferentes autores têm mostrado que a maioria das propriedades da distribuição normal continua valendo na classe das distribuições elípticas. Contudo, algumas propriedades, tais como a equivalência entre independência e covariância nula de variáveis aleatórias que têm distribuição conjunta normal, são próprias desta distribuição e permitem caracterizá-la dentro da classe das distribuições elípticas. A distribuição elíptica $n$-variada pode ser obtida misturando a distribuição uniforme sobre a superfície da $n$-esfera com uma distribuição radial (definida sobre $[0, \infty)$). Assim, cada distribuição elíptica ou, equivalentemente, sua versão padronizada, chamada comumente de esférica, é completamente caracterizada pela distribuição radial (veja Cambanis et al., 1981). Em particular, a distribuição elíptica (ou esférica) normal $n$-variada é caracterizada pela distribuição radial qui-quadrado com $n$ graus de liberdade. Similarmente, a distribuição elíptica $t$-Student $n$-variada com $\nu$ graus de liberdade é caracterizada pela distribuição radial de Fisher com $n$ e $\nu$ graus de liberdade.

A inferência nos modelos lineares sob a classe das distribuições elípticas tem sido objeto de estudo por muitos autores, trabalhos que podem ser encontrados, por exemplo, em Fang e Zhang (1990). No entanto, na literatura não encontramos um estudo minucioso dos modelos lineares com erros nas variáveis na classe das distribuições elípticas. Inferência nestes modelos é importante pelas suas aplicações em diferentes áreas do conhecimento, como por exemplo, ciências médicas ou ciências sociais, onde encontramos problemas envolvendo relações entre duas variáveis que podem descrever leis, em que certas variáveis estão funcionalmente relacionadas. Apesar dos constantes avanços tecnológicos terem tornado cada vez mais precisos os procedimentos de mensuração, não é realista supor que tais variáveis são medidas sem erros e o mais comum é não ter acesso aos seus verdadeiros valores. O modelo linear com erros nas variáveis mencionado acima é definido a seguir. Considere a relação linear,

$$y_i = \alpha + \beta x_i, \tag{1.1}$$

onde o interesse é, em geral, estimar o parâmetro $\beta$, mas nem $y$ nem $x$ são observados diretamente, ou seja, os valores observados são $Y$ e $X$, onde

$$\begin{cases} Y_i = y_i + e_i \\ X_i = x_i + u_i \end{cases} \tag{1.2}$$

com $i = 1, \ldots, n$. Na literatura, notamos que o modelo definido acima é apresentado sob normalidade seguindo três especificações: modelo ultraestrutural, modelo funcional e modelo estrutural. O modelo ultraestrutural foi introduzido por Dolby (1976) e assume que

$$e_i \sim N(0, \sigma_{ee}), \; u_i \sim N(0, \sigma_{uu}), \; x_i \sim N(\mu_{x_i}, \sigma_{xx}),$$

onde $e_1, \ldots, e_n$, $u_1, \ldots, u_n$, $x_1, \ldots, x_n$ são variáveis aleatórias independentes. Um estudo de inferência sobre o parâmetro (ou coeficiente angular) $\beta$ pode ser encontrado em Gleser (1985) e Cheng e Van Ness (1991).

O modelo funcional é obtido supondo que os valores não observados $x_1, \ldots, x_n$ são constantes (denominados parâmetros incidentais). Nesta situação, o número de parâmetros cresce com o número de observações e a função de verossimilhança não é limitada. Solari (1969) considerara suposições adicionais para possibilitar a inferência do parâmetro $\beta$.

O modelo estrutural é obtido considerando-se $x_1, \ldots, x_n$ como variáveis aleatórias com média comum $\mu_x$ e variância $\sigma_{xx}$. Também neste modelo não é possível encontrar uma solução de máxima verossimilhança para $\beta$, porém por um motivo diferente do descrito acima para o modelo funcional, pois neste caso o modelo não é identificável, no sentido de que dois diferentes vetores de parâmetros podem dar origem à mesma distribuição. Assim, é necessário o conhecimento adicional de algum parâmetro do modelo para possibilitar a estimação de $\beta$ (veja Fuller, 1987).

Em um contexto mais geral, o modelo ultraestrutural sobre a classe das distribuições elípticas é obtido supondo que o vetor $\mathbf{r}_i = (x_i, e_i, u_i)^\top$ tem uma distribuição elíptica com vetor de locação $\boldsymbol{\eta}_i = (\mu_{x_i}, 0, 0)^\top$ e matriz de dispersão $\boldsymbol{\Omega} = \mathrm{diag}(\sigma_{xx}, \sigma_{ee}, \sigma_{uu})$, onde a notação $\mathrm{diag}(a_1, \ldots, a_p)$ denota a matriz diagonal com elementos $a_1, \ldots, a_p$. Em particular, obtemos o modelo elíptico funcional, considerando que $\mu_{x_i} = x_i$, $i = 1, \ldots, n$ e o modelo estrutural considerando que $\mu_{x_i} = \mu_x$, $i = 1, \ldots, n$.

O modelo de regressão sobre a classe das distribuições elípticas, nas situações em que o número de parâmetros não varia com o número de observações é apresentado na literatura sob duas especificações diferentes. A primeira assume que a distribuição conjunta das variáveis de interesse é elíptica e no segundo enfoque as variáveis de interesse são independentes e tem distribuições marginais elípticas. No modelo elíptico estrutural com erro nas variáveis, os dois enfoques são denominados por:

(i) **Modelo elíptico dependente:** assume que os erros têm distribuição conjunta elíptica, de modo que eles podem ter covariância nula, sem serem independentes, a menos que a distribuição conjunta seja normal. Como pode ser visto, por exemplo, em Zellner (1976), Ghosh e Sinha (1980), Ullah e Walsh (1984) e Anderson, Fang e Hsu (1986), grande parte da inferência estatística baseada no modelo normal permanece válida para estes modelos elípticos dependentes, de modo que tais procedimentos são robustos com respeito a este tipo de não normalidade.

(ii) **Modelo elíptico independente:** supõe que os erros (ou observações) são independentes e têm distribuição marginal elíptica. Este enfoque é considerado, por exemplo, em Lange et al. (1989), Taylor (1992) e Kano, Berkane e Bentler (1993). Neste caso, em geral, a inferência deve ser baseada em resultados assintóticos. Contudo, a vantagem destes modelos elípticos independentes é que muitos deles, tais como o modelo $t$-Student, implicam em inferências robustas com respeito a observações aberrantes. Convém notar que sob normalidade ambos os enfoques são equivalentes. Isto mostra que a inferência estatística normal pode ser generalizada para cobrir uma classe mais ampla de modelos paramétricos que preservam a simetria do modelo normal.

Agora, suponhamos que no modelo definido em (1.1) e (1.2), $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_q)^\top$ e $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_q)^\top$ e a seqüência $X_1, \ldots, X_n$ são variáveis aleatórias unidimensionais. O modelo resultante que denominamos multi-unidimensional é usado na literatura no estudo de calibração comparativa, tanto no enfoque estrutural obtido supondo que as quantidades $x_1, \ldots, x_n$ são variáveis aleatórias com média $\mu_x$ e variância $\sigma_{xx}$, quanto no enfoque funcional obtido supondo que as quantidades $x_1, \ldots, x_n$ são constantes.

## 1.2 Definição dos objetivos

O objetivo deste trabalho é a inferência estatística nos modelos com erros nas variáveis sobre a classe das distribuições elípticas. No modelo ultraestrutural, a proposta é estender os resultados de Gleser (1985) e Cheng e Van Ness (1991) para a classe das distribuições elípticas, que inclui também os modelos estrutural e funcional elíptico.

No modelo estrutural, estudamos o comportamento assintótico do estimador de máxima verossimilhança e a eficiência relativa assintótica deste estimador com os estimadores conhecidos na literatura sob o modelo normal. O estudo cobre tanto o caso do modelo independente como do dependente, considerados, por exemplo, por Zellner (1976) e Lange et al. (1989). Além disso, estudamos inferência para os coeficientes angulares de duas ou mais populações estruturais normais, utilizando um método diferente daquele considerado em Wong (1991).

No modelo funcional multi-univariado, discutimos o comportamento assintótico do estimador de máxima verossimilhança, baseado no trabalho de Mak (1982). Assim, os objetivos específicos desta tese podem ser resumidos como segue.

(1) No modelo ultraestrutural, considerando condições adicionais sobre os parâmetros incidentais, estudamos a consistência e a normalidade assintótica de alguns estimadores do coeficiente angular, usando resultados assintóticos derivados para a matriz

de covariâncias amostrais $S_n$. Particularmente, no modelo estrutural discutimos a eficiência relativa assintótica do estimador de máxima verossimilhança com respeito aos estimadores estudados via a distribuição da matriz de covariâncias amostrais $S_n$.

(2) Estudar a consistência e a normalidade assintótica do estimador de máxima verossimilhança do parâmetro (vetor) estrutural do modelo funcional elíptico multi-univariado. Em particular, no caso univariado, estudar a eficiência relativa assintótica do estimador de máxima verossimilhança com respeito ao estimador de Mínimos Quadrados Generalizados.

(3) Discutir a inferência sobre os coeficientes angulares de duas ou mais populações estruturais sob normalidade, obtendo fatores de correção do tipo Bartlett através da avaliação direta do valor esperado da estatística da razão de verossimilhança sob a hipótese nula. Extensões são consideradas para a classe das distribuições elípticas.

## 1.3 Organização do trabalho

Este trabalho está organizado em quatro capítulos mais um apêndice.

No capítulo 2, começamos estudando a distribuição assintótica da matriz de covariâncias amostrais $S_n$ sob uma distribuição não especificada tendo quarto momento finito, onde o número de parâmetros do modelo cresce com o tamanho da amostra. Estes resultados são usados para o estudo do comportamento assintótico de alguns estimadores utilizados na literatura quando a normalidade é assumida. Sob o modelo elíptico estrutural, estudamos o comportamento assintótico do estimador de máxima verossimilhança. A eficiência assintótica é estudada com relação a estimadores obtidos por outros métodos. Inferência no modelo funcional para os estimadores de máxima verossimilhança será objeto de um estudo separado devido à complexidade do modelo.

No capítulo 3 estudamos inferência no modelo funcional quando o vetor de erros tem distribuição elíptica com vetor de locação zero e matriz de dispersão identidade. O estudo é baseado no trabalho de Mak (1982). Discutimos, com especial atenção, inferência no modelo que chamamos de multi-univariado, que é usado freqüentemente para a comparação de instrumentos ou métodos de medição, conhecido na literatura como de calibração comparativa. Apresentamos um estudo detalhado para o modelo $t$-Student.

No capítulo 4, discutimos inferência sobre os coeficientes angulares no modelo consistindo de duas ou mais populações estruturais. Primeiramente estendemos alguns resultados obtidos no modelo normal para os modelos elípticos. Três tipos de extensões do modelo sob normalidade para o modelo elíptico são consideradas. Posteriormente, sob o modelo normal tratamos o problema de testes de hipóteses, usando a estatística da razão de verossimilhança. Para obter a estatística da razão corrigida, o fator de correção de Bartlett é obtido avaliando-se diretamente o valor esperado da referida estatística sob a hipótese nula, que representa um método alternativo ao método de Lawley (1956). Os resultados obtidos são considerados também para a classe das distribuições elípticas.

# Capítulo 2

# O modelo ultraestrutural elíptico

## 2.1 Introdução

Neste capítulo estudaremos o modelo ultraestrutural sem réplicas introduzido por Dolby (1976) sobre a classe das distribuições elípticas, definido pelas relações lineares

$$\begin{cases} Y_k = y_k + e_k \\ X_k = x_k + u_k \\ y_k = \alpha + \beta x_k \ , \end{cases} \tag{2.1}$$

onde $(y_k, x_k)^\top$ são quantidades não observáveis e a estimação deve ser baseada nos dados observáveis $(Y_k, X_k)$, com $e_k \sim (0, \sigma_{ee})$, $u_k \sim (0, \sigma_{uu})$ os erros do modelo e $x_k \sim (\mu_{x_k}, \sigma_{xx})$, $k = 1, \ldots, n$, sendo todos independentes. A notação $(0, \sigma_{ee})$ significa que a variável $e_k$ segue uma distribuição não especificada, com média zero e variância $\sigma_{ee}$. A formulação acima é uma extensão dos modelos funcional e estrutural, pois, quando $\mu_{x_k} = \mu_x$, $k = 1, \ldots, n$, temos o modelo estrutural e quando $\sigma_{xx} = 0$, temos o modelo funcional. Neste último caso, $\mu_{x_k} = x_k$, $k = 1, \ldots, n$, são parâmetros incidentais e, portanto, sob este modelo o número de parâmetros cresce com o tamanho da amostra $n$. Um estudo das propriedades dos estimadores de máxima verossimilhança dos parâmetros estruturais é desenvolvido no Capítulo 3.

Nos modelos com erros nas variáveis (veja por exemplo Fuller, 1987, Gleser, 1985, e Cheng e Van Ness, 1991), a distribuição assintótica dos estimadores do coeficiente angular $\beta$ derivados na literatura, dependem da distribuição assintótica da matriz de covariâncias amostrais $\mathbf{S}_n$ de dimensão $2 \times 2$, das variáveis aleatórias $(Y_k, X_k)^\top$, cujas entradas são denotadas por $S_{YY} = \frac{1}{n}\sum_{k=1}^{n}(Y_k - \overline{Y})^2$, $S_{XY} = \frac{1}{n}\sum_{k=1}^{n}(X_k - \overline{X})(Y_k - \overline{Y})$ e $S_{XX} = \frac{1}{n}\sum_{k=1}^{n}(X_k - \overline{X})^2$. Esta distribuição é bem conhecida no caso em que os vetores $\mathbf{Z}_k = (Y_k, X_k)^\top$, $k = 1, \ldots, n$, são independentes e identicamente distribuídos, com momentos de quarta ordem finitos (veja Muirhead, 1982), como acontece com o modelo estrutural sob normalidade. No entanto, esta distribuição não pode ser usada no modelo funcional, devido à presença de parâmetros incidentais. Gleser (1985) e Cheng e Van Ness (1991) consideram a distribuição da matriz de covariâncias amostrais $\mathbf{S}_n$ sob o modelo ultraestrutural normal e mais recentemente Linder e Babu (1994), consideram que os erros são não correlacionados com terceiros momentos nulos, quartos momentos finitos e sob condições adicionais sobre os parâmetros incidentais para garantir a normalidade assintótica de $\mathbf{S}_n$.

Neste capítulo derivamos a distribuição assintótica conjunta do vetor de médias $\overline{\mathbf{Z}} = \frac{1}{n}\sum_{k=1}^{n}\mathbf{Z}_k$ e da matriz de covariâncias amostrais $\mathbf{S}_n = \frac{1}{n}\sum_{k=1}^{n}(\mathbf{Z}_k - \overline{\mathbf{Z}})(\mathbf{Z}_k - \overline{\mathbf{Z}})^\top$ de $n$ vetores $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ independentes de dimensão $p$ com vetores de médias $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_n$, respectivamente, e com matriz de covariâncias comum $\boldsymbol{\Sigma}$. Suposições tais como $\boldsymbol{\epsilon}_k = \mathbf{Z}_k - \boldsymbol{\mu}_k$, $k = 1, \ldots, n$, são independentes e identicamente distribuídas e condições assintóticas sobre os parâmetros $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_n$ são considerados. Mostramos também que o vetor de médias amostrais $\overline{\mathbf{Z}}_n$ e a matriz de covariâncias amostrais $\mathbf{S}_n$ são assintoticamente independentes quando a distribuição de $\boldsymbol{\epsilon}_k = \mathbf{Z}_k - \boldsymbol{\mu}_k$ tem terceiro momento nulo e quarto momento finito. Além disso, derivamos uma expressão geral para a matriz de covariâncias assintótica de $\sqrt{n}\mathbf{S}_n$. Uma expressão explícita é obtida para esta matriz no contexto das distribuições elípticas que é usada para obter a distribuição assintótica de alguns estimadores de $\beta$ no modelo ultraestrutural, generalizando e unificando os resultados de Gleser (1985) e Cheng e Van Ness (1994) para o contexto elíptico. Finalmente, os resultados são ilustrados no contexto da distribuição $t$-Student.

## 2.2 Notação e resultados preliminares

Nesta seção, apresentamos de forma resumida algumas definições, propriedades de operações com matrizes e propriedades dos cumulantes associadas à distribuição de um vetor aleatório.

**Definição 2.1** *1) Sejam* $\mathbf{A} = (a_{ij})$ *e* $\mathbf{B} = (b_{ij})$ *matrizes de dimensões* $m \times n$ *e* $r \times s$*, respectivamente, e seja* $\mathbf{A}_{\cdot j}$ *a* $j$*-ésima coluna da matriz* $\mathbf{A}$*,* $j = 1, \ldots, n$*. O vetor coluna* $\mathrm{Vec}(\mathbf{A})$ *de dimensão* $mn$ *é definido por*

$$\mathrm{Vec}(\mathbf{A}) = \begin{bmatrix} \mathbf{A}_{\cdot 1} \\ \mathbf{A}_{\cdot 2} \\ \vdots \\ \mathbf{A}_{\cdot n} \end{bmatrix}. \tag{2.2}$$

*2) O produto de Kronecker de* $\mathbf{A}$ *e* $\mathbf{B}$ *que denotamos por* $\mathbf{A} \otimes \mathbf{B}$ *é a matriz de dimensão* $mr \times ns$*, definida como*

$$\mathbf{A} \otimes \mathbf{B} = (a_{ij}\mathbf{B}) \tag{2.3}$$

*e a matriz* $\mathbf{A} * \mathbf{B}$ *de ordem* $mr \times ns$ *é definida por*

$$\mathbf{A} * \mathbf{B} = [\mathbf{A} \otimes \mathbf{B}_{\cdot 1}, \ldots, \mathbf{A} \otimes \mathbf{B}_{\cdot s}], \tag{2.4}$$

*onde* $\mathbf{B}_{\cdot j}$ *é a* $j$*-ésima coluna da matriz* $\mathbf{B}$*.*

O operador Vec de uma matriz simétrica $\mathbf{A}$ de dimensão $p \times p$ contém elementos que são iguais. O vetor que contém os $\frac{p}{2}(p+1)$ elementos distintos de $\mathbf{A}$ será denotado por $\mathrm{Vech}(\mathbf{A})$. Por exemplo, se $\mathbf{A} = \begin{pmatrix} a & b \\ b & d \end{pmatrix}$, então $\mathrm{Vec}(\mathbf{A}) = (a, b, b, d)^\top$ e $\mathrm{Vech}(\mathbf{A}) = (a, b, d)^\top$. Para uma matriz simétrica de dimensão $p \times p$, $\mathrm{Vec}(A)$ e $\mathrm{Vech}(A)$ são transformações lineares de um vetor com relação ao outro (veja Henderson e Searle, 1979), ou seja, existem matrizes $\mathbf{H}$ de dimensão $\frac{p}{2}(p+1) \times p^2$, e $\mathbf{G}$ de dimensão $p^2 \times \frac{p}{2}(p+1)$, tais que

$$\mathrm{Vech}(\mathbf{A}) = \mathbf{H}\mathrm{Vec}(\mathbf{A}) \quad \text{e} \quad \mathrm{Vec}(\mathbf{A}) = \mathbf{G}\mathrm{Vech}(\mathbf{A}). \tag{2.5}$$

Além disso,

$$\mathbf{HG} = \mathbf{I}_q, \quad q = \frac{p}{2}(p+1) \quad \text{e} \quad \mathbf{I}_{(p,p)}\mathbf{G} = \mathbf{G}, \tag{2.6}$$

onde $\mathbf{I}_{(p,p)}$ é a matriz de Vec-permutação. A matriz $\mathbf{G}$ em (2.5) é única. Contudo para transformar Vec($\mathbf{A}$) em Vech($\mathbf{A}$) pode existir mais de uma matriz $\mathbf{H}$ (veja Henderson e Searle, 1979). Pode-se mostrar que uma delas pode ser obtida de (2.6) que é a matriz inversa de Moore-Penrose, dada por

$$\mathbf{H} = (\mathbf{G}^\top \mathbf{G})^{-1}\mathbf{G}^\top. \tag{2.7}$$

Assim, por exemplo, para $p = 2$, temos que as matrizes $\mathbf{G}$ e $\mathbf{H}$ são dadas por

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{e} \quad \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{2.8}$$

respectivamente.

Seja $\mathbf{X} = (X_1, \ldots, X_m)^\top$ um vetor aleatório de dimensão $m$ com função característica $\phi(\mathbf{t})$, $\mathbf{t} \in \mathbb{R}^m$, e suponhamos por simplicidade que existem os momentos da distribuição deste vetor. Assim, se a função característica de $X_j$, $j = 1, \ldots, m$, é $\phi_j(t_j) = \phi(\mathbf{t})$, onde $\mathbf{t} = (0, \ldots, 0, t_j, \ldots, 0)^\top$ e os cumulantes de $X_j$ são os coeficientes $\kappa_k^j$ em

$$\log \phi_j(t_j) = \sum_{k=1}^{\infty} \kappa_k^j \frac{(it_j)^k}{k!} .$$

Por exemplo, os quatro primeiros cumulantes em termos dos momentos $\mu_k^j = E[X_j^k]$ de $X_j$ são dados por

$$\begin{aligned}
\kappa_1^j &= \mu_1^j, \\
\kappa_2^j &= \mu_2^j - (\mu_1^j)^2 = \mathrm{Var}[X_j], \\
\kappa_3^j &= \mu_3^j - 3\mu_1^j\mu_2^j + 2(\mu_1^j)^3, \\
\kappa_4^j &= \mu_4^j - 4\mu_1^j\mu_3^j - 3(\mu_2^j)^3 + 12\mu_2^j(\mu_1^j)^3 - 6(\mu_1^j)^4.
\end{aligned}$$

Os cumulantes mistos são definidos de forma similar, ou seja, se $\phi_{jk}(t_j, t_k)$ é a função característica conjunta de $X_j$ e $X_k$, $j, k = 1, \ldots, m$, os cumulantes de sua distribuição conjunta são os coeficientes $\kappa^{jk}_{r_1 r_2}$ em

$$\log \phi_{jk}(t_i, t_k) = \sum_{r_1, r_2 = 0}^{\infty} \kappa^{jk}_{r_1 r_2} \frac{(it_j)^{r_1}(it_k)^{r_2}}{r_1! r_2!},$$

onde, por exemplo, $\kappa^{jk}_{11} = \text{Cov}[X_j, X_k]$. Em geral, os cumulantes da distribuição conjunta de $X_1, \ldots, X_m$ são dados pelos coeficientes $\kappa^{1\ 2\cdots m}_{r_1 r_2 \cdots r_m}$ em

$$\log \phi_{\mathbf{X}}(\mathbf{t}) = \sum_{r_1, r_2, \cdots, r_m = 0} \kappa^{1\ 2\cdots m}_{r_1 r_2 \cdots r_m} \frac{(it_1)^{r_1}(it_2)^{r_2} \cdots (it_m)^{r_m}}{r_1! r_2! \cdots r_m!}.$$

**Lema 2.1** (Muirhead, 1982) *Seja* $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_p)^\top \sim (E[\boldsymbol{\epsilon}], \text{Var}[\boldsymbol{\epsilon}]) = (\mathbf{0}, \boldsymbol{\Sigma})$, *onde* $\boldsymbol{\Sigma} = (\sigma_{ij})$. *Suponhamos que* $E[\epsilon_i \epsilon_j \epsilon_k \epsilon_l] < \infty$, $i, j, k, l = 1, \ldots, p$. *Seja* $\mathbf{E} = (E_{ij}) = \boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top$ *e* $\boldsymbol{\Lambda} = \text{Var}[\text{Vec}(\mathbf{E})]$. *Então, a matriz* $\boldsymbol{\Lambda}$ *é determinada por*

$$\text{Cov}[E_{ij}, E_{kl}] = \kappa^{ijkl}_{1111} + \kappa^{ik}_{11}\kappa^{jl}_{11} + \kappa^{il}_{11}\kappa^{jk}_{11}, \tag{2.9}$$

*onde os coeficientes* $\kappa$*'s denotam os cumulantes da distribuição do vetor* $\boldsymbol{\epsilon}$. *Em particular,* $\kappa^{ij}_{11} = \text{Cov}[\epsilon_i, \epsilon_j] = \sigma_{ij}$, $\kappa^{iijk}_{1111} = \kappa^{ijk}_{211}$, $\kappa^{ii}_{11} = \kappa^{i}_{2} = \text{Var}[\epsilon_i] = \sigma_{ii}$, *etc.*

O seguinte resultado é uma versão do Lema 2.1 sob a classe das distribuições elípticas (veja a definição no Apêndice A).

**Lema 2.2** (Muirhead, 1982) *Suponha que* $\boldsymbol{\epsilon} \sim E\ell_p(\mathbf{0}, \boldsymbol{\Psi}; \phi)$, *com função característica* $\phi_{\boldsymbol{\epsilon}}(\mathbf{t}) = \phi(\mathbf{t}^\top \boldsymbol{\Psi} \mathbf{t})$, $\mathbf{t} \in \mathbb{R}^p$. *Então sob as condições do Lema 2.1,*

$$E[\boldsymbol{\epsilon}] = 0, \quad \text{Var}[\boldsymbol{\epsilon}] = \boldsymbol{\Sigma} = -2\phi'(0)\boldsymbol{\Psi},$$

$$\kappa^{ijkl}_{1111} = \kappa(\sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}), \tag{2.10}$$

*onde*

$$\kappa = \frac{\phi''(0)}{(\phi'(0))^2} - 1, \tag{2.11}$$

*conhecido como parâmetro de curtose.*

De (2.9) e (2.10), segue que se $\boldsymbol{\epsilon} \sim E\ell_p(\mathbf{0}, \boldsymbol{\Psi}; \phi)$ tem momento de quarta ordem finitos e $\mathrm{Var}[\boldsymbol{\epsilon}] = \Sigma = (\sigma_{ij})$, então a matriz $\boldsymbol{\Lambda} = \mathrm{Var}[\mathrm{Vec}(\mathbf{E})]$, com $\mathbf{E} = \boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top$, é determinada por

$$\mathrm{Cov}[E_{ij}, E_{kl}] = (\kappa + 1)(\sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}) + \kappa\sigma_{ij}\sigma_{kl}\ . \tag{2.12}$$

Se a distribuição de $\boldsymbol{\epsilon}$ tem densidade, então ela é da forma $|\boldsymbol{\Psi}|^{-\frac{1}{2}} f(\boldsymbol{\epsilon}^\top \boldsymbol{\Psi}^{-1}\boldsymbol{\epsilon})$, para alguma função não negativa $f(u)$, $u \geq 0$. Por exemplo, se $\boldsymbol{\epsilon} \sim E\ell_p(\mathbf{0}, \boldsymbol{\Psi}; f)$, com $f$ dada por

$$f(u) = k(p, \nu)\nu^{\nu/2}\{\nu + u\}^{-(\nu+p)/2}, \quad u \geq 0, \tag{2.13}$$

onde $k(p, \nu) = \Gamma\left[\frac{1}{2}(\nu + p)\right] / \pi^{p/2}\Gamma\left(\frac{1}{2}\nu\right)$, ou seja, $\boldsymbol{\epsilon} \sim t_p(\mathbf{0}, \boldsymbol{\Psi}; \nu)$, a distribuição $t$-Student $p$-variada com $\nu$ graus de liberdade, então (veja Muirhead, 1982),

$$\kappa = \frac{\nu - 2}{\nu - 4} - 1, \quad \nu > 4$$

e para o modelo normal $\kappa = 0$, que se obtém fazendo $\nu \to \infty$.

## 2.3 Propriedades assintóticas de $\mathbf{S}_n$

Nesta seção, apresentamos alguns resultados assintóticos para a matriz de covariâncias amostrais $\mathbf{S}_n$ sob a presença dos parâmetros incidentais, que serão de utilidade no estudo do modelo ultraestrutural.

Seja $\mathbf{Z}_1, \mathbf{Z}_2, \ldots$ uma seqüência de vetores aleatórios de dimensão $p$, com vetor de locação $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots$, respectivamente. Seja $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^\top$ e suponhamos que $\boldsymbol{\mu}_k = (\mu_{1k}, \mu_{2k}, \ldots, \mu_{pk})^\top$. Se $\Delta_{ik} = \mu_{ik} - \mu_k$, $i = 1, \ldots, p$, segue que

$$\boldsymbol{\mu}_k - \boldsymbol{\mu} = \begin{pmatrix} \Delta_{1k} \\ \Delta_{2k} \\ \vdots \\ \Delta_{pk} \end{pmatrix} \quad \text{e} \quad \overline{\mathbf{W}}_* = \frac{1}{n}\sum_{k=1}^{n} \mathbf{W}_k^*, \tag{2.14}$$

onde $\mathbf{W}_k^* = (\boldsymbol{\mu}_k - \boldsymbol{\mu})(\boldsymbol{\mu}_k - \boldsymbol{\mu})^\top$. Agora, vamos supor que a seqüência de vetores aleatórios $\mathbf{Z}_1, \mathbf{Z}_2, \ldots$ satisfaça as seguintes condições:

A1. $\epsilon_k = \mathbf{Z}_k - \boldsymbol{\mu}_k \overset{\text{iid}}{\sim} (\mathbf{0}, \boldsymbol{\Sigma})$, ou seja, vetor aleatório com vetor de locação $\mathbf{0}$ e matriz de dispersão $\boldsymbol{\Sigma}$;

A2. existe um vetor $\boldsymbol{\mu}$ (definido em (2.4)) de dimensão $p$ e uma matriz definida positiva $\boldsymbol{\Sigma}^*$ de dimensão $p \times p$ tais que, quando $n \to \infty$,

$$\sqrt{n}(\overline{\boldsymbol{\mu}} - \boldsymbol{\mu}) \to 0, \quad \sqrt{n}(\mathbf{S}^* - \boldsymbol{\Sigma}^*) \to 0 \quad \text{e} \quad \delta_n = \max_{1 \leq i \leq p} \max_{1 \leq k \leq n} \frac{\Delta_{ik}^2}{n} \to 0,$$

onde $\overline{\boldsymbol{\mu}} = \frac{1}{n} \sum_{k=1}^n \boldsymbol{\mu}_k$ e $\mathbf{S}^* = \frac{1}{n} \sum_{k=1}^n (\boldsymbol{\mu}_k - \overline{\boldsymbol{\mu}})(\boldsymbol{\mu}_k - \overline{\boldsymbol{\mu}})^\top$;

A3. $\text{Var}[\text{Vec}(\epsilon_1 \epsilon_1^\top)] = \boldsymbol{\Lambda} < \infty$, isto é, $\epsilon_1$ tem quarto momento finito;

A4. $E[\epsilon_1 \{\text{Vec}(\epsilon_1 \epsilon_1^\top)\}^\top] = \mathbf{0}$, que acontece, por exemplo, se a distribuição de $\epsilon_1$ é simétrica.

A seguir estudaremos o comportamento assintótico, da média e a matriz de covariâncias amostrais de $\overline{\mathbf{Z}}$ e $\mathbf{S}_n$, respectivamente, sob as suposições citadas acima.

**Lema 2.3** *Sob as suposições A1 e A2,*

$$i)\ \overline{\mathbf{Z}} \to \boldsymbol{\mu} \ q.c. \quad e \quad ii)\ \mathbf{S}_n \to \boldsymbol{\Sigma} + \boldsymbol{\Sigma}^* \ q.c.$$

**Prova:** i) A demonstração segue da relação

$$\overline{\mathbf{Z}} - \boldsymbol{\mu} = (\overline{\mathbf{Z}} - \overline{\boldsymbol{\mu}}) + (\overline{\boldsymbol{\mu}} - \boldsymbol{\mu})$$

e da lei forte dos grandes números.

ii) Note que podemos escrever

$$\mathbf{S}_n = \overline{\mathbf{W}} + \overline{\mathbf{W}}^* - (\overline{\mathbf{Z}} - \boldsymbol{\mu})(\overline{\mathbf{Z}} - \boldsymbol{\mu})^\top \tag{2.15}$$

onde $\overline{\mathbf{W}}^*$ é como em (2.14) e $\overline{\mathbf{W}} = \frac{1}{n} \sum_{k=1}^n \mathbf{W}_k$, com

$$\mathbf{W}_k = (\mathbf{Z}_k - \boldsymbol{\mu})(\mathbf{Z}_k - \boldsymbol{\mu})^\top = \epsilon_k \epsilon_k^\top + \epsilon_k (\boldsymbol{\mu}_k - \boldsymbol{\mu})^\top + (\boldsymbol{\mu}_k - \boldsymbol{\mu}) \epsilon_k^\top \tag{2.16}$$

e $\epsilon_k = \mathbf{Z}_k - \boldsymbol{\mu}_k$, $k = 1, \ldots, n$. Como $\overline{\mathbf{W}}_* = \mathbf{S}^* + (\overline{\mu} - \boldsymbol{\mu})(\overline{\mu} - \boldsymbol{\mu})^\top$, segue da condição A2 que

$$\lim \overline{\mathbf{W}}^* = \boldsymbol{\Sigma}^* \tag{2.17}$$

Por outro lado,

$$\lim \frac{1}{n} \sum_{k=1}^{n} \boldsymbol{\epsilon}_k (\boldsymbol{\mu}_k - \boldsymbol{\mu})^\top = \lim \frac{1}{n} \sum_{k=1}^{n} (\boldsymbol{\mu}_k - \boldsymbol{\mu}) \boldsymbol{\epsilon}_k^\top = \mathbf{0}, \qquad \text{q.c.}$$

pois da condição A2 e usando a fórmula da soma parcial de Abel (veja Gleser, 1981), temos que

$$\mathrm{Var}(\epsilon_{j1}) \sum_{k=1}^{\infty} \frac{\Delta_{ik}^2}{k^2} = \sum_{k=1}^{\infty} \mathrm{Var}(\Delta_{ik} \epsilon_{jk}) < \infty,$$

o que implica que $\lim \frac{1}{n} \sum_{k=1}^{n} \Delta_{ik} \epsilon_{jk} = 0$, $i, j = 1, \ldots, p$, ou seja,

$$\lim \frac{1}{n} \sum_{k=1}^{n} (\boldsymbol{\mu}_i - \boldsymbol{\mu}) \boldsymbol{\epsilon}_k^\top = 0 \text{ q.c.}$$

Similarmente, temos também que $\lim n^{-1} \sum_{k=1}^{n} \boldsymbol{\epsilon}_k (\boldsymbol{\mu}_k - \boldsymbol{\mu})^\top = 0$ q.c. Assim, temos de (2.16) que

$$\lim \overline{\mathbf{W}} = \boldsymbol{\Sigma}, \qquad \text{q.c.} \tag{2.18}$$

e a demonstração segue de (2.15), (2.17), (2.18) e do fato que $(\overline{\mathbf{Z}} - \boldsymbol{\mu})(\overline{\mathbf{Z}} - \boldsymbol{\mu})^\top \to 0$ q.c. $\square$

O resultado seguinte será de grande utilidade na demonstração dos principais resultados deste capítulo.

**Teorema 2.1** *Sob as condições A1 - A3,*

$$\sqrt{n}(\overline{\mathbf{Z}} - \boldsymbol{\mu}) \xrightarrow{\mathrm{d}} N_p(\mathbf{0}, \boldsymbol{\Sigma}) \quad e \quad \sqrt{n}\mathrm{Vec}(\mathbf{S}_n - \boldsymbol{\Sigma} - \boldsymbol{\Sigma}^*) \xrightarrow{\mathrm{d}} N_{p^2}(\mathbf{0}, \boldsymbol{\Lambda} + \boldsymbol{\Lambda}^*),$$

*onde* $\boldsymbol{\Lambda} = \mathrm{Var}[\mathrm{Vec}(\boldsymbol{\epsilon}_1 \boldsymbol{\epsilon}_1^\top)]$, *cujos elementos são determinados por (2.9) no Lema 2.1 e*

$$\boldsymbol{\Lambda}^* = \boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}^* + \boldsymbol{\Sigma}^* \otimes \boldsymbol{\Sigma} + \boldsymbol{\Sigma} * \boldsymbol{\Sigma}^* + \boldsymbol{\Sigma}^* * \boldsymbol{\Sigma}\ , \tag{2.19}$$

*onde a notação "$\xrightarrow{\mathrm{d}}$" indica a convergência em distribuição.*

**Prova:** De (2.15), temos que

$$\sqrt{n}\begin{pmatrix}\text{Vec}(\mathbf{S}_n - \boldsymbol{\Sigma} - \boldsymbol{\Sigma}^*) \\ \overline{\mathbf{Z}} - \boldsymbol{\mu}\end{pmatrix} = \sqrt{n}\begin{pmatrix}\text{Vec}(\overline{\mathbf{W}} - \boldsymbol{\Sigma}) \\ \bar{\boldsymbol{\epsilon}}\end{pmatrix} + \sqrt{n}\begin{pmatrix}\text{Vec}(\overline{\mathbf{W}}^* - \boldsymbol{\Sigma}) \\ \bar{\boldsymbol{\mu}} - \boldsymbol{\mu}\end{pmatrix}$$

$$-\sqrt{n}\begin{pmatrix}\text{Vec}((\overline{\mathbf{Z}} - \boldsymbol{\mu})(\overline{\mathbf{Z}} - \boldsymbol{\mu})^\top) \\ 0\end{pmatrix}.$$

Da condição A2 e de i) do Lema 2.3, temos que

$$\sqrt{n}\begin{pmatrix}\text{Vec}(\mathbf{S}_n - \boldsymbol{\Sigma} - \boldsymbol{\Sigma}^*) \\ \overline{\mathbf{Z}} - \boldsymbol{\mu}\end{pmatrix} = \sqrt{n}\begin{pmatrix}\text{Vec}(\overline{\mathbf{W}} - \boldsymbol{\Sigma}) \\ \bar{\boldsymbol{\epsilon}}\end{pmatrix} + O_p(1). \tag{2.20}$$

Agora, seja

$$\mathbf{Y}_k = \begin{pmatrix}\text{Vec}(\mathbf{W}_k - \boldsymbol{\Sigma}) \\ \boldsymbol{\epsilon}_k\end{pmatrix},$$

$k = 1, \ldots, n$. Então, de A1 e A3, segue que $\mathbf{Y}_1, \mathbf{Y}_2, \ldots$ são vetores independentes de dimensão $p^2 + p$, com $E[\mathbf{Y}_k] = \mathbf{0}$ e $\text{Var}[\mathbf{Y}_k] = \boldsymbol{\Delta}_k < \infty$, onde

$$\boldsymbol{\Delta}_k = \begin{pmatrix}\boldsymbol{\Lambda}_k & \boldsymbol{\Psi}_k \\ \boldsymbol{\Psi}_k & \boldsymbol{\Sigma}\end{pmatrix},$$

com $\boldsymbol{\Psi}_k = \text{Cov}[\boldsymbol{\epsilon}_k, \text{Vec}(\mathbf{W}_k)]$ e $\boldsymbol{\Lambda}_k = \text{Var}[\text{Vec}(\mathbf{W}_k)]$. De A1 temos que

$$\begin{aligned}\boldsymbol{\Psi}_k &= E[\boldsymbol{\epsilon}_1[\text{Vec}(\boldsymbol{\epsilon}_1\boldsymbol{\epsilon}_1^\top)\}^\top] + E[\boldsymbol{\epsilon}_1\{\text{Vec}(\boldsymbol{\epsilon}_1(\boldsymbol{\mu}_k - \boldsymbol{\mu})^\top\}^\top] + E[\boldsymbol{\epsilon}_1\{\text{Vec}((\boldsymbol{\mu}_k - \boldsymbol{\mu})\boldsymbol{\epsilon}_1^\top)\}^\top] \\ &= \boldsymbol{\Psi} + E[\boldsymbol{\epsilon}_1\{\text{Vec}(\boldsymbol{\epsilon}_1(\boldsymbol{\mu}_k - \boldsymbol{\mu})^\top)\}^\top] + E[\boldsymbol{\epsilon}_1\{\text{Vec}(\boldsymbol{\mu}_k - \boldsymbol{\mu})\boldsymbol{\epsilon}_1^\top)\}^\top],\end{aligned}$$

onde $\boldsymbol{\Psi} = E[\boldsymbol{\epsilon}_1\{\text{Vec}(\boldsymbol{\epsilon}_1\boldsymbol{\epsilon}_1^\top)\}^\top]$. Assim, pela condição A2, segue que

$$\overline{\boldsymbol{\Psi}} = \frac{1}{n}\sum_{k=1}^{n}\boldsymbol{\Psi}_k \to \boldsymbol{\Psi}.$$

Similarmente,

$$\boldsymbol{\Lambda}_k = \text{Var}[\text{Vec}(\mathbf{W}_k)] = \boldsymbol{\Lambda} + \boldsymbol{\Lambda}_{1k} - \boldsymbol{\Lambda}_{2k},$$

onde

$$\begin{aligned}\boldsymbol{\Lambda} &= \text{Var}[\text{Vec}(\boldsymbol{\epsilon}_1\boldsymbol{\epsilon}_1^\top)], \\ \boldsymbol{\Lambda}_{1k} &= \text{Var}[\text{Vec}(\boldsymbol{\epsilon}_1(\boldsymbol{\mu}_k - \boldsymbol{\mu})^\top + (\boldsymbol{\mu}_k - \boldsymbol{\mu})\boldsymbol{\epsilon}_1^\top)]\end{aligned}$$

e

$$
\begin{aligned}
\Lambda_{2k} &= \mathrm{Cov}[\mathrm{Vec}(\boldsymbol{\epsilon}_1\boldsymbol{\epsilon}_1^{\top}), \mathrm{Vec}(\boldsymbol{\epsilon}_1(\boldsymbol{\mu}_k-\boldsymbol{\mu})^{\top}+(\boldsymbol{\mu}_k-\boldsymbol{\mu})\boldsymbol{\epsilon}_1^{\top})] \\
&\quad + \mathrm{Cov}[\mathrm{Vec}(\boldsymbol{\epsilon}_1(\boldsymbol{\mu}_k-\boldsymbol{\mu})^{\top}+(\boldsymbol{\mu}_k-\boldsymbol{\mu})\boldsymbol{\epsilon}_1^{\top}, \mathrm{Vec}(\boldsymbol{\epsilon}_1\boldsymbol{\epsilon}_1^{\top})] \,.
\end{aligned}
$$

Assim, sob a condição A2, temos que

$$
\frac{1}{n}\sum_{k=1}^{n}\Lambda_{2k}\to 0 \quad \text{e} \quad \frac{1}{n}\sum_{k=1}^{n}\Lambda_{1k}\to\Lambda^{*}=\Sigma\otimes\Sigma^{*}+\Sigma^{*}\otimes\Sigma+\Sigma*\Sigma^{*}+\Sigma^{*}*\Sigma,
$$

e conseqüentemente

$$
\overline{\Lambda}=\frac{1}{n}\sum_{k=1}^{n}\Lambda_k\to\Lambda+\Lambda^{*},
$$

o que implica que

$$
\overline{\Delta}=\frac{1}{n}\sum_{k=1}^{n}\Delta_k=\begin{pmatrix}\overline{\Lambda} & \overline{\Psi}\\ \overline{\Psi} & \Sigma\end{pmatrix}\to\Delta=\begin{pmatrix}\Lambda+\Lambda^{*} & \Psi\\ \Psi & \Sigma\end{pmatrix}.
$$

Para provar a normalidade assintótica de $\sqrt{n}\overline{\mathbf{Y}}$, só resta demonstrar que a seqüência $\mathbf{Y}_1, \mathbf{Y}_2, \ldots$ satisfaz a condição de Lindeberg, isto é, $\forall\varepsilon>0$,

$$
\frac{1}{n}\sum_{k=1}^{n}E\left[\|\mathbf{Y}_k\|^2 I_{\{\|\mathbf{Y}_k\|>\varepsilon\sqrt{n}\}}\right]\to 0.
$$

Com efeito, seja

$$
\mathbf{Y}_k=\begin{pmatrix}\mathbf{X}_k\\ \boldsymbol{\epsilon}_k\end{pmatrix},
$$

onde

$$
\mathbf{X}_k=\mathrm{Vec}(\mathbf{W}_k-\mathbf{\Sigma})=\mathrm{Vec}(\boldsymbol{\epsilon}_k\boldsymbol{\epsilon}_k^{\top}-\mathbf{\Sigma})+\mathrm{Vec}((\boldsymbol{\mu}_k-\boldsymbol{\mu})\boldsymbol{\epsilon}_k^{\top}+\boldsymbol{\epsilon}_k(\boldsymbol{\mu}_k-\boldsymbol{\mu})^{\top}).
$$

Então,

$$
\|\mathbf{Y}_k\|^2=\|\mathbf{X}_k\|^2+\|\boldsymbol{\epsilon}_k\|^2 \quad \text{e} \quad \|\mathbf{Y}_k\|\le\|\mathbf{X}_k\|+\|\boldsymbol{\epsilon}_k\|.
$$

Assim,

$$
\frac{1}{n}\sum_{k=1}^{n}E\left[\|\mathbf{Y}_k\|^2 I_{\{\|\mathbf{Y}_k\|>\varepsilon\sqrt{n}\}}\right]\le A_{n\varepsilon}+B_{n\varepsilon}+C_{n\varepsilon}+D_{n\varepsilon}, \tag{2.21}
$$

onde

$$
\begin{aligned}
A_{n\varepsilon} &= \frac{1}{n}\sum_{k=1}^{n} E\left[\|\mathbf{X}_k\|^2 I_{\{\|\mathbf{X}_k\|>\frac{\varepsilon}{2}\sqrt{n}\}}\right], \\
B_{n\varepsilon} &= \frac{1}{n}\sum_{k=1}^{n} E\left[\|\mathbf{X}_k\|^2 I_{\{\|\boldsymbol{\epsilon}_k\|>\frac{\varepsilon}{2}\sqrt{n}\}}\right], \\
C_{n\varepsilon} &= \frac{1}{n}\sum_{k=1}^{n} E\left[\|\boldsymbol{\epsilon}_k\|^2 I_{\{\|\mathbf{X}_k\|>\frac{\varepsilon}{2}\sqrt{n}\}}\right] \qquad \text{e} \\
D_{n\varepsilon} &= \frac{1}{n}\sum_{k=1}^{n} E\left[\|\boldsymbol{\epsilon}_k\|^2 I_{\{\|\boldsymbol{\epsilon}_k\|>\frac{\varepsilon}{2}\sqrt{n}\}}\right].
\end{aligned}
$$

Claramente, temos de A1 que $D_{n\varepsilon} \to 0$. A seguir, mostraremos que $A_{n\varepsilon} \to 0$. Com efeito, como

$$
(\boldsymbol{\mu}_x - \boldsymbol{\mu})\boldsymbol{\epsilon}_k^\top + \boldsymbol{\epsilon}_k(\boldsymbol{\mu}_k - \boldsymbol{\mu})^\top = (\mathbf{B}_1\Delta_{1k} + \cdots + \mathbf{B}_p\Delta_{pk})\boldsymbol{\epsilon}_k,
$$

onde $\mathbf{B}_i$, $i = 1, \ldots, p$, são matrizes constantes de dimensão $p^2 \times p$ e $\Delta_{ik}$, $i = 1, \ldots, p$ são como em (2.14), temos que

$$
\begin{aligned}
\|X_k\| &\leq \|\mathrm{Vec}(\boldsymbol{\epsilon}_k\boldsymbol{\epsilon}_k^\top - \boldsymbol{\Sigma})\| + \|[\mathbf{I}_p \otimes (\mathbf{B}_1\Delta_{1k} + \cdots + \mathbf{B}_p\Delta_{pk})]\boldsymbol{\epsilon}_k\| \\
&\leq \|\mathrm{Vec}(\boldsymbol{\epsilon}_k\boldsymbol{\epsilon}_k^\top - \boldsymbol{\Sigma})\| + \sqrt{p}\|\mathbf{B}_1\Delta_{1k} + \cdots + \mathbf{B}_p\Delta_{pk}\| \, \|\boldsymbol{\epsilon}_k\| \\
&\leq \|\mathrm{Vec}(\boldsymbol{\epsilon}_k\boldsymbol{\epsilon}_k^\top - \boldsymbol{\Sigma})\| + \sqrt{p}M\sum_{i=1}^{p}|\Delta_{ik}|\|\boldsymbol{\epsilon}_k\|
\end{aligned}
$$

e

$$
\|X_k\|^2 \leq 4\left\{\|\mathrm{Vec}(\boldsymbol{\epsilon}_k\boldsymbol{\epsilon}_k^\top - \boldsymbol{\Sigma})\| + pM^2c_p\sum_{i=1}^{p}\Delta_{ik}^2\|\boldsymbol{\epsilon}_k\|^2\right\},
$$

com $M = \max_{1\leq i\leq p}\|B_i\|$, o que implica que

$$
A_{n\varepsilon} \leq A_{n\varepsilon}^{(1)} + A_{n\varepsilon}^{(2)} + A_{n\varepsilon}^{(3)} + A_{n\varepsilon}^{(4)}, \tag{2.22}
$$

onde

$$
\begin{aligned}
A_{n\varepsilon}^{(1)} &= 4E\left[\|\mathrm{Vec}(\boldsymbol{\epsilon}_1\boldsymbol{\epsilon}_1^\top - \boldsymbol{\Sigma})\|^2 I_{\|\mathrm{Vec}(\boldsymbol{\epsilon}_1\boldsymbol{\epsilon}_1^\top - \boldsymbol{\Sigma})\|>\frac{\varepsilon}{4}\sqrt{n}\}}\right], \\
A_{n\varepsilon}^{(2)} &= 4E\left[\|\mathrm{Vec}(\boldsymbol{\epsilon}_1\boldsymbol{\epsilon}_1^\top - \boldsymbol{\Sigma})\|^2 I_{\left\{\|\boldsymbol{\epsilon}_1\|>\frac{\varepsilon}{4\sqrt{p}M\delta_n}\right\}}\right],
\end{aligned}
$$

$$
\begin{aligned}
A_{n\varepsilon}^{(3)} &= \frac{4pc_pM^2}{n}\sum_{k=1}^{n}\sum_{i=1}^{p}\Delta_{ik}^2 E\left[\|\boldsymbol{\epsilon}_1\|^2 I_{\|\mathrm{Vec}(\boldsymbol{\epsilon}_1\boldsymbol{\epsilon}_1^\top-\boldsymbol{\Sigma})\|>\frac{\varepsilon}{4}\sqrt{n}\}}\right] \qquad \text{e}\\
A_{n\varepsilon}^{(4)} &= \frac{4pc_pM^2}{n}\sum_{k=1}^{n}\sum_{i=1}^{p}\Delta_{ik}^2 E\left[\|\boldsymbol{\epsilon}_1\|^2 I_{\|\boldsymbol{\epsilon}_1\|^2>\frac{\varepsilon}{4\sqrt{p}M\delta_n}\}}\right].
\end{aligned}
$$

Usando o Teorema da Convergência Dominada e a condição $\lim\delta_n = 0$ (onde $\delta_n$ é como em A2), segue que as esperanças nas igualdades acima convergem para zero e, portanto $\lim A_{n\varepsilon}^{(l)} = 0$, $l = 1,2,3,4$, pois $\lim n^{-1}\sum_{k=1}^{n}\sum_{i=1}^{p}\Delta_{ik}^2 < \infty$. Logo, de (2.22) temos que $\lim A_{n\varepsilon} = 0$. Similarmente, mostra-se que $\lim B_{n\varepsilon} = \lim C_{n\varepsilon} = \lim D_{n\varepsilon} = 0$. Assim, de (JÁ MOSTROU) (2.21), temos que a condição de Lindberg é satisfeita e, conseqüentemente,

$$\sqrt{n}\mathbf{Y} \xrightarrow{\mathrm{d}} N_{p^2+p}(\mathbf{0},\boldsymbol{\Delta}) \tag{2.23}$$

e da relação em (2.20) e do Teorema de Slutsky, segue que

$$\sqrt{n}(\overline{\mathbf{Z}}-\boldsymbol{\mu}) \xrightarrow{\mathrm{d}} N_p(\mathbf{0},\boldsymbol{\Sigma})$$

e

$$\sqrt{n}\mathrm{Vec}(\mathbf{S}_n-\boldsymbol{\Sigma}-\boldsymbol{\Sigma}^*) \xrightarrow{\mathrm{d}} N_{p^2}(\mathbf{0},\boldsymbol{\Lambda}+\boldsymbol{\Lambda}^*).$$

□

Como $\mathbf{S}_n$ é uma matriz simétrica, então o vetor $\mathrm{Vec}(\mathbf{S}_n)$ contém elementos que são iguais, de modo que a distribuição assintótica de $\sqrt{n}\mathrm{Vec}(\mathbf{S}_n)$ é singular. Contudo, se considerarmos os $(p^2+p)/2$ elementos distintos de $\mathbf{S}_n$, ou seja, o vetor $\sqrt{n}\mathrm{Vech}(\mathbf{S}_n)$, temos que sua distribuição assintótica será não singular, como veremos a seguir.

**Corolário 2.1** *Sob as condições do Teorema 2.1,*

$$\sqrt{n}\mathrm{Vech}(\mathbf{S}_n-\boldsymbol{\Sigma}-\boldsymbol{\Sigma}^*) \xrightarrow{\mathrm{d}} N_q(\mathbf{0},\boldsymbol{\nabla}+\boldsymbol{\nabla}^*), \tag{2.24}$$

*onde* $\boldsymbol{\nabla} = \mathbf{H}\boldsymbol{\Lambda}\mathbf{H}^\top$, $\boldsymbol{\nabla}^* = \mathbf{H}\boldsymbol{\Lambda}^*\mathbf{H}^\top$, $q = p(p+1)/2$ *e* $\mathbf{H}$ *é como em (2.5). Além disso, se a condição A4 é satisfeita,* $\overline{\mathbf{Z}}$ *e* $\mathbf{S}_n$ *são assintoticamente independentes.*

**Prova:** De (2.5), temos que existe uma matriz $\mathbf{H}$ de dimensão $p(p+1)/2 \times p^2$ tal que

$$\text{Vech}(\mathbf{S}_n - \Sigma - \Sigma^*) = \mathbf{H}\text{Vec}(\mathbf{S}_n - \Sigma - \Sigma^*).$$

A demonstração de (2.24) segue da igualdade acima, juntamente com o Teorema 2.2. 2.1
Agora, como $\text{Var}[\epsilon_1\{\text{Vec}(\epsilon_1\epsilon_1^\top)\}^\top] = 0$, temos que a matriz $\Psi = 0$. Assim, de (2.23) temos que $\overline{\mathbf{Z}}$ e $\mathbf{S}_n$ são assintoticamente independentes. □

## 2.4 O modelo elíptico ultraestrutural

Consideremos o modelo (2.1), que pode ser representado na forma

$$\mathbf{Z}_k = \mathbf{a} + \mathbf{B}\mathbf{r}_k, \tag{2.25}$$

onde

$$\mathbf{Z} = \begin{pmatrix} Y_k \\ X_k \end{pmatrix}, \quad \mathbf{r}_k = \begin{pmatrix} x_k \\ e_k \\ u_k \end{pmatrix}, \quad \mathbf{a} = \begin{pmatrix} \alpha \\ 0 \end{pmatrix} \quad \text{e} \quad \mathbf{B} = \begin{pmatrix} \beta & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix},$$

$k = 1, \ldots, n$, satisfazendo as seguintes condições:

S1. $\mathbf{r}_k \overset{\text{ind}}{\sim} E\ell_3(\boldsymbol{\eta}_k, \Psi; \phi)$, com $E[\mathbf{r}_k] = \boldsymbol{\eta}_k = (\mu_{x_k}, 0, 0)^\top$ e $\text{Var}[\mathbf{r}_k] = \boldsymbol{\Omega} = -2\phi'(0)\Psi = \text{diag}(\sigma_{xx}, \sigma_{ee}, \sigma_{uu})$ finitos.

S2. Existem $\mu_x$ e $\sigma^*_{xx} > 0$ finitos tais que

$$\sqrt{n}(\overline{\mu}_x - \mu_x) \to 0, \quad \sqrt{n}(S^*_{xx} - \sigma^*_{xx}) \to 0 \quad \text{e} \quad \max_{1 \leq k \leq n} \frac{(\mu_{x_k} - \mu_x)^2}{n} \to 0,$$

onde

$$\overline{\mu}_x = \frac{1}{n}\sum_{k=1}^{n} \mu_{x_k} \quad \text{e} \quad S^*_{xx} = \frac{1}{n}\sum_{k=1}^{n} (\mu_{x_k} - \overline{\mu}_x)^2.$$

S3. $\mathbf{r}_1 - \boldsymbol{\eta}_1 = (x_1 - \mu_{x_1}, e_1, u_1)^\top \sim E\ell_3(\mathbf{0}, \Psi; \phi)$, tem momentos de quarta ordem finitos.

**Lema 2.4** *A seqüência $\mathbf{Z}_1, \mathbf{Z}_2, \ldots$, definida em (2.25), sob as condições S1 - S3, satisfazem as condições A1 - A4.*

**Prova:** A1: Pelas propriedades das distribuições elípticas, temos de S1 que $\mathbf{Z}_k = \mathrm{a}+\mathbf{B}\mathbf{r}_k \overset{\text{iid}}{\sim} E\ell_2(\mathrm{a}+\mathbf{B}\boldsymbol{\eta}_k, \mathbf{B}\boldsymbol{\Psi}\mathbf{B}^\top; \phi)$, com $E[\mathbf{Z}_k] = \boldsymbol{\mu}_k$ e $\mathrm{Var}[\mathbf{Z}_k] = \boldsymbol{\Sigma}$, dados por

$$\boldsymbol{\mu}_k = \mathrm{a}+\mathbf{B}\boldsymbol{\eta}_k = \begin{pmatrix} \alpha+\beta\mu_{x_k} \\ \mu_{x_k} \end{pmatrix} \quad \text{e} \quad \boldsymbol{\Sigma} = \mathbf{B}\boldsymbol{\Omega}\mathbf{B}^\top = \begin{pmatrix} \beta^2\sigma_{xx}+\sigma_{ee} & \beta\sigma_{xx} \\ \beta\sigma_{xx} & \sigma_{xx}+\sigma_{uu} \end{pmatrix},$$

respectivamente. Assim, temos que $\boldsymbol{\epsilon}_k = \mathbf{Z}_k - \boldsymbol{\mu}_k \overset{\text{iid}}{\sim} E\ell_2(\mathbf{0}, \mathbf{B}\boldsymbol{\Psi}\mathbf{B}^\top; \phi)$, com $E[\boldsymbol{\epsilon}_k] = \mathbf{0}$ e $\mathrm{Var}[\boldsymbol{\epsilon}_k] = \boldsymbol{\Sigma}$.

A2: Note que $\overline{\boldsymbol{\mu}} = \frac{1}{n}\sum_{k=1}^n \boldsymbol{\mu}_k = (\alpha+\beta\overline{\mu}_x, \overline{\mu}_x)^\top$ e $\mathbf{S}^* = \frac{1}{n}\sum_{k=1}^n (\boldsymbol{\mu}_k - \overline{\boldsymbol{\mu}})(\boldsymbol{\mu}_k - \overline{\boldsymbol{\mu}})^\top = S_{xx}^*\mathbf{b}\mathbf{b}^\top$, onde $\mathbf{b} = (\beta, 1)^\top$. Agora, sejam

$$\boldsymbol{\mu} = (\alpha+\beta\mu_x, \mu_x)^\top \quad \text{e} \quad \boldsymbol{\Sigma}^* = \sigma_{xx}^*\mathbf{b}\mathbf{b}^\top. \tag{2.26}$$

Então,

$$\sqrt{n}(\overline{\boldsymbol{\mu}} - \boldsymbol{\mu}) = \sqrt{n}(\overline{\mu}_x - \mu_x)\mathrm{b} \to 0 \quad \text{e} \quad \sqrt{n}(\mathbf{S}^* - \boldsymbol{\Sigma}^*) = \sqrt{n}(S_{xx}^* - \sigma_{xx}^*)\mathbf{b}\mathbf{b}^\top \to 0.$$

Por outro lado, para $\Delta_{1k} = (\alpha+\beta\mu_{x_k} - \alpha+\beta\mu_x) = \beta(\mu_{x_k} - \mu_x)$ e $\Delta_{2k} = \mu_{x_k} - \mu_x$, temos que $\max_{1\le k\le n} \frac{\Delta_{ik}^2}{n} \to 0$, $i = 1, 2$, desde que $\max_{1\le k\le n} \frac{(\mu_{x_k}-\mu_x)^2}{n} \to 0$.

A prova de A3 e A4 segue diretamente de S3 e da simetria da família das distribuições elípticas, respectivamente. □

Sejam $\overline{\mathbf{Z}}$ e $\mathbf{S}_n$ a média e matriz de covariâncias amostrais de $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$. Então, pelo Lema 2.4, as condições S1 – S3 são satisfeitas. Logo, usando o Teorema 2.1, temos o seguinte resultado.

**Teorema 2.2** *Sob o modelo ultraestrutural satisfazendo as condições S1 – S3,*

*i)* $\overline{\mathbf{Z}} \to \mathrm{a}+\mathbf{b}\mu_x$ *q.c. e* $\mathbf{S}_n \to \boldsymbol{\Sigma} + \sigma_{xx}^*\mathbf{b}\mathbf{b}^\top$ *q.c.;*

*ii)* $\sqrt{n}(\overline{\mathbf{Z}} - \mathrm{a} - \mathbf{b}\mu_x) \to N_2(\mathbf{0}, \boldsymbol{\Sigma})$ *e* $\sqrt{n}\mathrm{Vech}(\mathbf{S}_n - \boldsymbol{\Sigma} - \sigma_{xx}^*\mathbf{b}\mathbf{b}^*) \to N_3(\mathbf{0}, \nabla + \nabla^*)$*, onde*

$$\nabla + \nabla^* = \begin{pmatrix} (3\kappa+2)\sigma_{YY}^2 & (3\kappa+2)\sigma_{YY}\sigma_{YX} & \kappa\sigma_{YY}\sigma_{XX} + 2(\kappa+1)\sigma_{YX}^2 \\ & (\kappa+1)\sigma_{YY}\sigma_{XX} + (2\kappa+1)\sigma_{YX}^2 & (3\kappa+2)\sigma_{YX}\sigma_{XX} \\ & & (3\kappa+2)\sigma_{XX}^2 \end{pmatrix}$$

$$+\sigma_{xx}^* \begin{pmatrix} 4\beta^2\sigma_{YY} & 2\beta\sigma_{YY} + 2\beta^2\sigma_{YX} & 4\beta\sigma_{YX} \\ & \sigma_{YY} + 2\beta\sigma_{YX} + \beta^2\sigma_{XX} & 2\sigma_{YX} + 2\beta\sigma_{XX} \\ & & 4\sigma_{XX} \end{pmatrix}$$

**Prova:** Seja $\boldsymbol{\epsilon}_k = \mathbf{Z}_k - \boldsymbol{\mu}_k$, $k = 1, \ldots, n$. Então $E[\boldsymbol{\epsilon}_1] = 0$, $\mathrm{Var}[\boldsymbol{\epsilon}_1] = \boldsymbol{\Sigma} = (\sigma_{ij})$. Agora, seja $\mathbf{E} = \boldsymbol{\epsilon}_1\boldsymbol{\epsilon}_1^\top = (E_{ij})$ e

$$\boldsymbol{\Lambda} = \mathrm{Var}[\mathrm{Vec}(\mathbf{E})] = \begin{pmatrix} \mathrm{Var}[E_{11}] & \mathrm{Cov}[E_{11}, E_{12}] & \mathrm{Cov}[E_{11}, E_{21}] & \mathrm{Cov}[E_{11}, E_{22}] \\ & \mathrm{Var}[E_{12}] & \mathrm{Cov}[E_{12}, E_{21}] & \mathrm{Cov}[E_{12}, E_{22}] \\ & & \mathrm{Var}[E_{21}] & \mathrm{Cov}[E_{21}, E_{22}] \\ & & & \mathrm{Var}[E_{22}] \end{pmatrix}.$$

Assim, usando o Lema 2.2 para determinar os elementos $\mathrm{Cov}[E_{ij}, E_{lm}]$, temos, após algumas manipulações algébricas, que

$$\boldsymbol{\Lambda} = \begin{pmatrix} (3\kappa+2)\sigma_{YY}^2 & (3\kappa+2)\sigma_{YY}\sigma_{YX} & (3\kappa+2)\sigma_{YY}\sigma_{YX} & \kappa\sigma_{YY}\sigma_{XX} + c_1 \\ & (\kappa+1)\sigma_{YY}\sigma_{XX} + c_2 & (\kappa+1)\sigma_{YY}\sigma_{XX} + c_2 & (3\kappa+2)\sigma_{XY}\sigma_{XX} \\ & & (\kappa+1)\sigma_{YY}\sigma_{XX} + c_2 & (3\kappa+2)\sigma_{XY}\sigma_{XX} \\ & & & (3\kappa+2)\sigma_{XX}^2 \end{pmatrix}, \tag{2.27}$$

onde $c_1 = 2(\kappa+1)\sigma_{YX}^2$, $c_2 = (2\kappa+1)\sigma_{YX}^2$, $\sigma_{11} = \sigma_{YY} = \mathrm{Var}[Y_i] = \beta^2\sigma_{xx} + \sigma_{ee}$, $\sigma_{12} = \sigma_{YX} = \mathrm{Cov}[Y_i, X_i] = \beta\sigma_{xx}$, $\sigma_{22} = \sigma_{XX} = \mathrm{Var}[X_i] = \sigma_{xx} + \sigma_{uu}$ e $\kappa$ é o coeficiente de curtose definido em (2.11). Agora, como $\boldsymbol{\Sigma}^* = \sigma_{xx}^*\mathbf{b}\mathbf{b}^\top$, $\mathbf{b} = (\beta, 1)^\top$, temos de (2.19) que

$$\boldsymbol{\Lambda}^* = \sigma_{xx}^* \begin{pmatrix} 4\beta^2\sigma_{YY} & 2\beta\sigma_{YY} + 2\beta^2\sigma_{YX} & 2\beta\sigma_{YY} + 2\beta^2\sigma_{YX} & 4\beta\sigma_{YX} \\ & \sigma_{YY} + 2\beta\sigma_{YX} + \beta^2\sigma_{XX} & \sigma_{YY} + 2\beta\sigma_{YX} + \beta^2\sigma_{XX} & 2\sigma_{YX} + 2\beta\sigma_{XX} \\ & & \sigma_{YY} + 2\beta\sigma_{YX} + \beta^2\sigma_{XX} & 2\sigma_{YX} + 2\beta\sigma_{XX} \\ & & & 4\sigma_{XX} \end{pmatrix}.$$

Assim, pelo Teorema 2.1, temos que

$$\sqrt{n}(\overline{\mathbf{Z}} - \mathbf{a} - \mathbf{b}\mu_x) \xrightarrow{\mathrm{d}} N_2(\mathbf{0}, \boldsymbol{\Sigma}) \quad \text{e} \quad \sqrt{n}\mathrm{Vec}(\mathbf{S}_n - \boldsymbol{\Sigma} - \sigma_{xx}^*\mathbf{b}\mathbf{b}^\top) \xrightarrow{\mathrm{d}} N_4(\mathbf{0}, \boldsymbol{\Lambda} + \boldsymbol{\Lambda}^*),$$

cuja distribuição assintótica é singular, pois $\boldsymbol{\Lambda} + \boldsymbol{\Lambda}^*$ é simétrica. Assim, para obter uma distribuição não singular, vamos considerar o $\mathrm{Vech}(\mathbf{S}_n)$. Com efeito, usando a relação

(2.5) com $\mathbf{H}$ dado em (2.8), temos que

$$\sqrt{n}\text{Vech}(\text{S}_n - \Sigma - \sigma_{xx}^*\mathbf{b}\mathbf{b}^\top) \to N_2(0, \nabla + \nabla^*),$$

onde

$$\nabla + \nabla^* = \mathbf{H}(\Lambda + \Lambda^*)\mathbf{H}^\top.$$

**Observação 2.1** Note que os resultados anteriores são independentes de condições adicionais sobre as variâncias dos erros. Contudo, para possibilitar a inferência, no modelo ultraestrutural normal (veja Gleser, 1985, e Cheng e Van Ness, 1991), condições adicionais sobre as variâncias dos erros são requeridas para tornar factível o problema de estimação dos parâmetros do modelo, especificamente, do parâmetro $\beta$, que é de maior interesse. Similarmente ao modelo ultraestrutural normal, vamos estudar o modelo definido em (2.25) sob as seguintes suposições: i) $\lambda_e = \sigma_{ee}/\sigma_{uu}$ conhecido e ii) $\sigma_{uu}$ (ou $\sigma_{ee}$) conhecido, consideradas por Gleser (1985) e Cheng e Van Ness (1991), respectivamente, sob normalidade. Em termos de $\lambda_e = \sigma_{ee}/\sigma_{uu}$, $\lambda_x = \sigma_{xx}/\sigma_{uu}$ e $\lambda_x^* = \sigma_{xx}^*/\sigma_{uu}$, temos que

$$\sigma_{YY} = \sigma_{uu}(\beta^2\lambda_x + \lambda_e), \quad \sigma_{YX} = \beta\sigma_{uu}\lambda_x \quad \text{e} \quad \sigma_{XX} = \sigma_{uu}(\lambda_x + 1).$$

Assim,

$$\Sigma = \sigma_{uu}\begin{pmatrix} \beta^2\lambda_x + \lambda_e & \beta\lambda_x \\ \beta\lambda_x & \lambda_x + 1 \end{pmatrix}, \quad \Sigma^* = \sigma_{uu}\begin{pmatrix} \beta^2\lambda_x^* & \beta\lambda_x^* \\ \beta\lambda_x^* & \lambda_x^* \end{pmatrix}$$

e a matriz de covariância assintótica de $\sqrt{n}\text{Vech}(\text{S}_n)$ dada por $\nabla_H = \nabla + \nabla^*$ (veja Teorema 2.2), pode ser escrita como

$$\begin{aligned} \nabla_H &= \sigma_{uu}^2 \begin{pmatrix} (3\kappa+2)(\beta^2\lambda_x+\lambda_e)^2 & (3\kappa+2)(\beta^2\lambda_x+\lambda_e)\beta\lambda_x & \kappa(\beta^2\lambda_x+\lambda_e)(\lambda_x+1)+c_1 \\ & (\kappa+1)(\beta^2\lambda_x+\lambda_e)(\lambda_x+1)+c_2 & (3\kappa+2)\beta\lambda_x(\lambda_x+1) \\ & & (3\kappa+2)(\lambda_x+1)^2 \end{pmatrix} \\ &\quad + \lambda_x^*\sigma_{uu}^2 \begin{pmatrix} 4\beta^2(\beta^2\lambda_x+\lambda_e) & 2\beta(\beta^2\lambda_x+\lambda_e) & 4\beta^2\lambda_x \\ & \beta^2(4\lambda_x+1)+\lambda_e & 2\beta(2\lambda_x+1) \\ & & 4(\lambda_x+1) \end{pmatrix}, \end{aligned} \tag{2.28}$$

onde $c_1$ e $c_2$ são como em (2.27).

Da observação anterior, temos dois casos especiais do modelo definido em (2.25), a saber:

**Modelo estrutural:** Consideremos $\mu_{x_k} = \mu_x$, $k = 1, \ldots, n$. Então, $\sigma^*_{xx} = 0$ e $\mathbf{\Lambda}^* = 0$ (veja Fuller, 1987). Assim, em particular, a matriz de covariâncias assintótica de $\sqrt{n}\text{Vech}(\mathbf{S}_n)$ sob normalidade ($\kappa = 0$) é dada por

$$\nabla = \sigma^2_{uu} \begin{pmatrix} 2(\beta^2\lambda_x + \lambda_e)^2 & 2(\beta^2\lambda_x + \lambda_e)\beta\lambda_x & 2\beta^2\lambda_x^2 \\ & (\beta^2\lambda_x + \lambda_e)(\lambda_x + 1) + \beta^2\lambda_x^2 & 2\beta\lambda_x(\lambda_x + 1) \\ & & 2(\lambda_x + 1)^2 \end{pmatrix} \tag{2.29}$$

**Modelo funcional:** O modelo ultraestrutural se reduz ao modelo funcional quando $\mu_{x_k} = x_k$, para todo $k$; logo, segue que $\lambda_x = 0$. Neste caso, a matriz de covariâncias assintótica do vetor $\sqrt{n}\text{Vech}(\mathbf{S}_n)$ sob o modelo elíptico é dado por

$$\nabla_H = \sigma^2_{uu} \begin{pmatrix} (3\kappa + 2)\lambda_e^2 & 0 & \kappa\lambda_e \\ & (\kappa + 1)\lambda_e & 0 \\ & & 3\kappa + 2 \end{pmatrix} + \lambda^*_x\sigma^2_{uu} \begin{pmatrix} 4\beta^2\lambda_e & 2\beta\lambda_e & 0 \\ & \beta^2 + \lambda_e & 2\beta \\ & & 4 \end{pmatrix}. \tag{2.30}$$

Além disso, sob normalidade,

$$\Lambda_H = \sigma^2_{uu} \begin{pmatrix} 2\lambda_e^2 & 0 & \\ & \lambda_e & 0 \\ & & 2 \end{pmatrix} + \lambda^*_x\sigma^2_{uu} \begin{pmatrix} 4\beta^2\lambda_e & 2\beta\lambda_e & 0 \\ & \beta^2 + \lambda_e & 2\beta \\ & & 4 \end{pmatrix}. \tag{2.31}$$

Note que a expressão (2.31) coincide com o resultado dado em Gleser (1985) com a notação $\sigma_{uu} = \theta$, $\lambda_e = \lambda$, $\lambda_x = \nu$, $\sigma^*_{xx} = \nu_*$ e $\lambda^*_x = \theta\nu$. Observamos que na matriz $\mathbf{\Lambda}$ dada na equação (9) daquele artigo, o termo $2\tau^2 - \nu_*^2$ deve ser substituído por $2(\tau^2 - \nu_*^2)$.

### 2.4.1 Razão de variâncias conhecida

Nesta seção, estudaremos o comportamento assintótico do estimador de $\beta$ sob a condição $\lambda_e = \sigma_{ee}/\sigma_{uu}$ conhecido. Assumindo que os vetores aleatórios $\mathbf{r}_1, \ldots, \mathbf{r}_n$ são tais que $E[\mathbf{r}_k] = \boldsymbol{\eta}_k$ e $\text{Var}[\mathbf{r}_k] = \text{diag}(\sigma_{xx}, \sigma_{ee}, \sigma_{uu}) = \sigma_{uu}\text{diag}(\lambda_x, \lambda_e, 1)$, Gleser (1985) mostra que, para $\lambda_e$ conhecido, o estimador de Mínimos Quadrados Generalizado é obtido minimizando a função

$$Q_G(\alpha,\beta,\boldsymbol{\mu}_*^\top,\lambda_x,\sigma_{uu})=\sum_{k=1}^{n}\boldsymbol{\epsilon}_k^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}_k, \tag{2.32}$$

com $\boldsymbol{\epsilon}_k=\mathbf{Z}_k-\boldsymbol{\mu}_k=\mathbf{B}(\mathbf{r}_k-\boldsymbol{\eta}_k)$, $\boldsymbol{\mu}_k=\mathbf{a}+\mathbf{B}\boldsymbol{\eta}_k$, $\boldsymbol{\Sigma}=\mathbf{B}\boldsymbol{\Omega}\mathbf{B}^\top$ e $\boldsymbol{\mu}^*=(\mu_{x_1},\ldots,\mu_{x_n})^\top$. Algumas manipulações algébricas mostram que este estimador é dado por

$$\widehat{\beta}_{GLS}=\frac{S_{YY}-\lambda_e S_{XX}+\{(S_{YY}-\lambda_e S_{XX})^2+4\lambda_e S_{YX}^2\}^{1/2}}{2S_{YX}}. \tag{2.33}$$

Note que $\widehat{\beta}_{GLS}=g(\mathrm{Vech}(\mathbf{S}_n))$, onde $g$ é dado por

$$g(x,y,z)=\frac{x-\lambda_e z+\{(x-\lambda_e z)^2+4\lambda_e y^2\}^{1/2}}{2y},$$

que é contínua para $y\neq 0$. Assim, desde que a seqüência $\mu_{x_1},\mu_{x_2},\ldots$ satisfaz a condição S2 e $\sigma_{xx}+\sigma_{xx}^*=\sigma_{uu}\lambda_x+\sigma_{uu}\lambda_x^*>0$, segue que

$$\widehat{\beta}_{GLS}=g(\mathrm{Vech}(\mathbf{S}_n))\xrightarrow{\text{q.c.}}g(\mathrm{Vech}(\boldsymbol{\Sigma}+\sigma_{xx}^*\mathbf{b}\mathbf{b}^\top))=\beta,$$

isto é, $\widehat{\beta}_{GLS}$ é um estimador consistente de $\beta$. A seguir mostraremos que o estimador $\sqrt{n}\widehat{\beta}_{GLS}$ é assintoticamente normal na classe das distribuições elípticas.

**Teorema 2.3** *Considere o modelo ultraestrutural definido em (2.25) e suponha que $\lambda_e$ é conhecido e as condições S1 – S3 são satisfeitas. Se $\sigma_{xx}+\sigma_{xx}^*=\sigma_{uu}\lambda_x+\sigma_{uu}\lambda_x^*>0$, então*

$$\sqrt{n}(\widehat{\beta}_{GLS}-\beta)\xrightarrow{\text{d}}N(0,\Delta_{GLS}), \tag{2.34}$$

*onde*

$$\begin{aligned}\Delta_{GLS}&=(\kappa+1)\beta^2\frac{\sigma_{YY}\sigma_{XX}-\sigma_{YX}^2}{(\sigma_{YX}+\sigma_{YX}^*)^2}+\beta^2\sigma_{xx}^*\frac{\sigma_{YY}+\beta^2\sigma_{XX}-2\beta\sigma_{YX}}{(\sigma_{YX}+\sigma_{YX}^*)^2}\\&=\frac{(\kappa+1)(\beta^2\lambda_x+\lambda_e\lambda_x+\lambda_e)+\lambda_x^*(\beta^2+\lambda_e)}{(\lambda_x+\lambda_x^*)^2}\end{aligned} \tag{2.35}$$

**Prova:** Sob as condições S1 – S3, temos pelo Teorema 2.2 que $\sqrt{n}\mathrm{Vech}(\mathbf{S}_n-\boldsymbol{\Sigma}-\sigma_{xx}^*\mathbf{b}\mathbf{b}^\top)\xrightarrow{\text{d}}N(\mathbf{0},\boldsymbol{\nabla}_H)$. Como $\widehat{\beta}_{GLS}=g(\mathrm{Vech}(\mathbf{S}_n))$, então pelo método delta, segue que

$$\sqrt{n}(\widehat{\beta}_{GLS}-\beta)\xrightarrow{\text{d}}N(0,\mathbf{d}^\top\boldsymbol{\nabla}_H\mathbf{d}),$$

onde

$$\mathbf{d} = \frac{1}{(\sigma_{YX} + \sigma^*_{YX})(\beta^2 + \lambda_e)}(\beta^2, -\beta(\beta^2 - \lambda_e), -\lambda_e\beta^2)^\top$$

e

$$\mathbf{d}^\top \nabla_H \mathbf{d} = \frac{(\kappa+1)\beta^2(\sigma_{YY}\sigma_{XX} - \sigma^2_{YX}) + \sigma^*_{xx}\beta^2(\sigma_{YY} + \beta^2\sigma_{XX} - 2\beta\sigma_{YX})}{(\sigma_{YX} + \sigma^*_{YX})^2}.$$

A segunda parte da igualdade em (2.35) segue considerando $\nabla_H$ como em (2.28) na expressão $\mathbf{d}^\top\nabla_H\mathbf{d}$.

**Observação 2.2** Sob o modelo elíptico, temos que o estimador $\widehat{\beta}_{GLS}$ é assintoticamente independente de $\overline{\mathbf{Z}}$. Além disso, sob a normalidade do modelo ($\kappa = 0$), temos que

$$\Delta_{GLS} = \frac{\beta^2\lambda_x + \lambda_e\lambda_x + \lambda_e + \lambda^*_x(\beta^2 + \lambda_e)}{(\lambda_x + \lambda^*_x)^2},$$

como é dado em Gleser (1985), que mostra que sob o modelo ultraestrutural normal com $\lambda_e$ conhecido, $\widehat{\beta}_{GLS}$ é também o estimador de máxima verossimilhança de $\beta$.

Observemos também que sob o modelo estrutural ($\lambda^*_x = 0$), a variância assintótica de $\sqrt{n}\widehat{\beta}_{GLS}$ pode ser expressa como

$$\Delta_{GLS} = (\kappa+1)\beta^2\frac{(1-\rho^2_{YX})}{\rho^2_{YX}}, \tag{2.36}$$

onde $\rho_{YX} = \mathrm{Cov}[Y_k, X_k]/(\mathrm{Var}[Y_k]\mathrm{Var}[X_k])^{1/2}$.

### 2.4.2 Uma das variâncias conhecida

Nesta seção, vamos considerar que a variância $\sigma_{uu}$ é conhecida (resultados similares podem ser obtidos quando a variância $\sigma_{ee}$ é conhecida).

Seja a seqüência $\mathbf{Z}_1, \ldots, \mathbf{Z}_k$ como em (2.25). Similarmente à seção anterior, vamos supor também para efeito de inferência, que as condições S1 – S3 são satisfeitas e conseqüentemente a seqüência $\mathbf{Z}_1, \mathbf{Z}_2, \ldots$ satisfaz as condições A1 – A4 sobre o contexto das distribuições elípticas. Nesta seção, estudaremos as propriedades assintóticas do estimador

$$\widehat{\beta}_M = \frac{S_{YX}}{S_{XX} - \sigma_{uu}}, \tag{2.37}$$

que é um estimador consistente sob normalidade, isto é, quando $\mathbf{r}_k$ é normalmente distribuído, $k = 1, \ldots, n$, como é considerado em Cheng and Van Ness (1991). Nakamura (1990) obtém uma função de escore corrigida sob a suposição de que $e_k$ em (2.25) é normalmente distribuído e $\sigma_{uu}$ é conhecido. A suposição de normalidade não é requerida para $u_k$.

**Teorema 2.4** *Considere o modelo ultraestrutural definido em (2.25), supondo $\sigma_{uu}$ conhecido e as condições S1 - S3 satisfeitas. Se $\sigma_{xx} + \sigma^*_{xx} > 0$, então*

$$\sqrt{n}(\widehat{\beta}_M - \beta) \xrightarrow{\text{d}} N(0, \Delta_M), \tag{2.38}$$

*onde*

$$\begin{aligned} \Delta_M &= \frac{(\kappa+1)(\sigma_{XX}\sigma_{YY} - \sigma^2_{YX}) + (3\kappa+2)(\beta\sigma_{XX} - \sigma_{XY})^2}{(\sigma_{xx} + \sigma^*_{xx})^2} \\ &\quad + \frac{\sigma^*_{xx}(\beta^2\sigma_{XX} - 2\beta\sigma_{XY} + \sigma_{YY})}{(\sigma_{xx} + \sigma^*_{xx})^2} \\ &= \frac{(\kappa+1)(\beta^2\lambda_x + \lambda_e\lambda_x + \lambda_e) + (3\kappa+2)\beta^2 + (\beta^2 + \lambda_e)\lambda^*_x}{(\lambda_x + \lambda^*_x)^2} \end{aligned} \tag{2.39}$$

**Prova:** Considerando que $\widehat{\beta}_M = g(\text{Vech}(\mathbf{S}_n))$, onde $g(x, y, z) = y/(z - \sigma_{uu})$, $z \neq \sigma_{uu}$, segue do Teorema 2.2 que

$$\widehat{\beta}_M \xrightarrow{\text{q.c.}} g(\text{Vech}(\boldsymbol{\Sigma} + \sigma^*_{xx}\mathbf{b}\mathbf{b}^\top)) = \beta\ .$$

Agora, considerando o Teorema 2.1 e usando o método delta, obtemos a distribuição assintótica de $\widehat{\beta}_M$, dada por

$$\sqrt{n}(\widehat{\beta}_M - \beta) \xrightarrow{\text{d}} N(0, \mathbf{d}^\top \nabla_H \mathbf{d}),$$

onde

$$\mathrm{d} = \frac{1}{\sigma_{XX} + \sigma^*_{XX} - \sigma_{uu}}(0, 1, -\beta)^\top.$$

Como $\sigma^*_{XX} = \sigma^*_{xx}$ e $\sigma_{XX} = \sigma_{xx} + \sigma_{uu}$, segue que

$$\begin{aligned} \mathrm{d}^\top \nabla_H \mathrm{d} &= \frac{(\kappa+1)(\sigma_{XX}\sigma_{YY} - \sigma^2_{XY}) + (3\kappa+2)(\beta\sigma_{XX} - \sigma_{XY})^2}{(\sigma_{xx} + \sigma^*_{xx})^2} \\ &\quad + \frac{\sigma^*_{xx}(\beta^2\sigma_{XX} - 2\beta\sigma_{XY} + \sigma_{YY})}{(\sigma_{xx} + \sigma^*_{xx})^2}. \end{aligned}$$

A prova da segunda igualdade em (2.39) é similar à prova do Teorema 2.3. □

Note que sob o modelo estrutural independente, segue de (2.24), com $\lambda^*_x = 0$ que a variância assintótica de $\sqrt{n}\widehat{\beta}_M$ pode ser escrita como

$$\begin{aligned} \Delta_M &= \frac{(\kappa+1)(\sigma_{XX}\sigma_{YY} - \sigma^2_{XY}) + (3\kappa+2)(\beta\sigma_{XX} - \sigma_{XY})^2}{\sigma^2_{xx}} \\ &= (\kappa+1)\beta^2\frac{(1-\rho^2_{YX})}{\rho^2_{YX}} + (3\kappa+2)\frac{\beta^2}{\lambda^2_x}, \end{aligned} \tag{2.40}$$

onde $\lambda_x = \sigma_{xx}/\sigma_{uu}$. Além disso, sob normalidade, $\kappa = 0$, a variância assintótica se reduz a

$$\Delta_M = \beta^2\frac{(1-\rho^2_{XX})}{\rho^2_{XX}} + \frac{2\beta^2}{\lambda^2_x},$$

que coincide com a variância assintótica dada em Fuller (1987).

Por outro lado, sob o modelo ultraestrutural normal ($\kappa = 0$), a variância assintótica é dada por

$$\Delta_M = \frac{\sigma_{XX}\sigma_{YY} - \sigma^2_{XY} + 2(\beta\sigma_{XX} - \sigma_{XY})^2 + \sigma^*_{xx}(\beta^2\sigma_{XX} - 2\beta\sigma_{XY} + \sigma_{YY})}{(\sigma_{xx} + \sigma^*_{xx})^2},$$

que coincide com a variância assintótica dada em Cheng e Van Ness (1991).

## 2.5 O modelo estrutural elíptico

Nesta seção, estudaremos a distribuição assintótica do estimador de máxima verossimilhança do parâmetro $\beta$ do modelo definido em (2.25), considerando os modelos estruturais elípticos independente e dependente. Além disso, estudaremos a eficiência relativa assintótica do estimador de máxima verossimilhança sob o modelo estrutural elíptico independente, com os estimadores de $\beta$ considerados na seção anterior.

Sob o modelo estrutural independente, o estimador de máxima verossimilhança de $\beta$ não tem uma forma fechada e um método iterativo é requerido. O algoritmo EM pode ser usado, como é considerado, por exemplo, em Lange et al. (1989) e Arellano-Valle (1994). Como neste modelo o número de parâmetros não varia com o tamanho da amostra, a variância assintótica dos estimadores dos parâmetros é obtida invertendo-se a matriz de informação de Fisher.

O modelo estrutural elíptico independente obtém-se considerando que a seqüência de vetores aleatórios $\mathbf{r}_1, \ldots, \mathbf{r}_n$ definida em (2.25) é independente e identicamente distribuída com $\mathbf{r}_1 \sim E\ell_3(\boldsymbol{\eta}, \boldsymbol{\Psi}; \phi)$, onde $\boldsymbol{\eta} = \boldsymbol{\eta}_k$, $k = 1, \ldots$, isto é, $\mu_{x_k} = \mu_x$, $k = 1, \ldots, n$, e $\boldsymbol{\Psi}$ como em S1 da seção anterior. Nesta seção, vamos supor que a distribuição de $\mathbf{r}_1$ tem densidade, e pela definição de distribuição elíptica, a função densidade pode ser escrita como

$$|\boldsymbol{\Psi}|^{-1/2} f((\mathbf{r}_1 - \boldsymbol{\eta})^\top \boldsymbol{\Psi}^{-1} (\mathbf{r}_1 - \boldsymbol{\eta})),$$

onde $f(u) \geq 0$, $u \geq 0$, isto é, $\mathbf{r}_1 \sim E\ell_3(\boldsymbol{\eta}, \boldsymbol{\Psi}; f)$. Assim, a seqüência $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ definida em (2.25) são vetores iid, com $\mathbf{Z}_1 \sim E\ell_2(\boldsymbol{\mu}, \mathbf{V}; f)$, cuja densidade é dada por

$$|\mathbf{V}|^{-1/2} f((\mathbf{Z}_1 - \boldsymbol{\mu})^\top \mathbf{V}^{-1} (\mathbf{Z}_1 - \boldsymbol{\mu})),$$

onde $\boldsymbol{\mu} = \mathbf{a} + \mathbf{B}\boldsymbol{\eta}$ e $\mathbf{V} = \mathbf{B}\boldsymbol{\Psi}\mathbf{B}^\top$, com $\mathbf{a}$ e $\mathbf{B}$ como em (2.25). Para uma amostra de tamanho $n$, o logaritmo da função de verossimilhança é dado por

$$L(\boldsymbol{\theta}) = -\frac{n}{2} \log |\mathbf{V}| + \sum_{k=1}^{n} \log f(\|\mathbf{T}_k\|^2), \tag{2.41}$$

com

$$\boldsymbol{\theta} = (\alpha, \mu_x, \sigma_{xx}, \sigma_{ee}, \sigma_{uu}, \beta)^\top \quad \text{e} \quad \mathbf{T}_k = \mathbf{V}^{-1/2}(\mathbf{Z}_k - \boldsymbol{\mu}), \tag{2.42}$$

$k = 1, \ldots, n$. Sob as condições acima, o modelo não é identificável, como é apontado em Arellano-Valle (1994). Para tomar o modelo identificável, condições adicionais são necessárias. Nesta seção, vamos considerar as seguintes suposições,

$$\lambda_e = \frac{\sigma_{ee}}{\sigma_{uu}}, \text{ conhecido ou } \quad \lambda_x = \frac{\sigma_{xx}}{\sigma_{uu}} \text{ conhecido ou } \quad \sigma_{uu}(\sigma_{ee}) \text{ conhecido.}$$

Um estudo detalhado do modelo estrutural elíptico com $\lambda_e$ ou $\lambda_x$ conhecido é considerado em Arellano-Valle (1994). Por outro lado, com a suposição de que $\sigma_{ee}$ (ou $\sigma_{uu}$) é conhecida, temos o estudo desenvolvido por Cheng e Van Ness (1991) sob normalidade. Considerando $\lambda_e$ ou $\lambda_x$ conhecido, o vetor $\boldsymbol{\theta}$ definido em (2.42) se reduz a

$$\boldsymbol{\theta} = \begin{cases} (\alpha, \mu_x, \sigma_{ee}, \sigma_{uu}, \beta)^\top & , \text{ se } \lambda_x \text{ é conhecido,} \\ (\alpha, \mu_x, \sigma_{xx}, \sigma_{uu}, \beta)^\top & , \text{ se } \lambda_e \text{ é conhecido.} \end{cases} \tag{2.43}$$

Assim, $\boldsymbol{\mu} = \mu(\boldsymbol{\theta})$, $\mathbf{V} = V(\boldsymbol{\theta})$ e a relação

$$\boldsymbol{\theta} \to (\mu(\boldsymbol{\theta}), V(\boldsymbol{\theta}))$$

é um a um, desta forma, $\boldsymbol{\theta}$ é identificável (veja Arellano-Valle, 1994). Sob normalidade, a condição $\lambda_e$ conhecido é considerada em Wong (1989) e a suposição $\lambda_x$ conhecido é considerada em Bolfarine e Cordani (1993). Não é difícil verificar que esta última condição é equivalente ao conhecimento de $k_x = \frac{\lambda_x}{\lambda_x+1}$, conhecido na literatura como coeficiente de atenuação ou confiabilidade. Tal situação é comum em algumas áreas como, por exemplo, em economia, psicologia, sociologia, (veja Fuller, 1987). Sob as duas primeiras condições de identificabilidade, segue que a matriz $\mathbf{V}$ pode ser escrita como

$$\mathbf{V} = \begin{cases} \begin{pmatrix} \lambda_x\beta^2\sigma_{uu} + \sigma_{ee} & \lambda_x\beta\sigma_{uu} \\ \lambda_x\beta\sigma_{uu} & (\lambda_x + 1)\sigma_{uu} \end{pmatrix} & , \text{ se } \lambda_x \text{ é conhecido,} \\ \begin{pmatrix} \beta^2\sigma_{xx} + \lambda_e\sigma_{uu} & \beta\sigma_{xx} \\ \beta\sigma_{xx} & \sigma_{xx} + \sigma_{uu} \end{pmatrix} & , \text{ se } \lambda_e \text{ é conhecido,} \end{cases} \tag{2.44}$$

com determinante dado por

$$|\mathbf{V}| = \begin{cases} [\lambda_x\beta^2\sigma_{uu} + (\lambda_x + 1)\sigma_{ee}]\sigma_{uu} & , \text{ se } \lambda_x \text{ é conhecido,} \\ [\lambda_e\sigma_{uu} + (\beta^2 + \lambda_e)\sigma_{xx}]\sigma_{uu} & , \text{ se } \lambda_e \text{ é conhecido.} \end{cases} \tag{2.45}$$

### 2.5.1 A matriz de informação de Fisher

Seja $\mathbf{K}(\boldsymbol{\theta}) = (K_{ij})$ a matriz de informação do modelo estrutural elíptico definido em (2.25), onde $\boldsymbol{\theta}$ é como em (2.43), com o logaritmo da função de verossimilhança dado em (2.41). Assim, temos que

$$\frac{\partial L}{\partial \theta_i} = -\frac{n}{2}\mathrm{tr}\left(\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_i}\right) + \sum_{k=1}^{n} W_f(\|\mathbf{T}_k\|^2)\frac{\partial}{\partial \theta_i}\|\mathbf{T}_k\|^2,$$

onde $W_f(u) = f'(u)/f(u)$, $u \geq 0$, e $\mathbf{T}_k$ é como em (2.42) e

$$\frac{\partial}{\partial \theta_i}\|\mathbf{T}_k\|^2 = -2\frac{\partial \boldsymbol{\mu}^\top}{\partial \theta_i}\mathbf{V}^{-1/2}\mathbf{T}_k - \mathbf{T}_k^\top \mathbf{V}^{-1/2}\frac{\partial \mathbf{V}}{\partial \theta_i}\mathbf{V}^{-1/2}\mathbf{T}_k\ .$$

Como $\mathbf{T}_k \sim E\ell_2(\mathbf{0}, \mathbf{I}_2; f)$, então, usando a simetria da distribuição $\mathbf{T}_k$ em conjunto com (A.13) e (A.14) do Apêndice A, segue, após algumas manipulações algébricas (ver Arellano-Valle, 1994) que

$$\begin{aligned} K_{ij} \;=\; & \frac{4n}{p}a_f(2,1)\frac{\partial \boldsymbol{\mu}^\top}{\partial \theta_i}\mathbf{V}^{-1}\frac{\partial \boldsymbol{\mu}}{\partial \theta_j} \\ & +\frac{np}{2}\left\{\frac{1}{p(p+1)}a_f(2,2) - \frac{1}{4}\right\}\mathrm{tr}\left(\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_i}\right)\mathrm{tr}\left(\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_j}\right) \\ & +\frac{2n}{p(p+2)}a_f(2,2)\mathrm{tr}\left(\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_i}\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_j}\right), \end{aligned} \tag{2.46}$$

onde

$$p = 2 \quad \text{e} \quad a_f(r,s) = E[(W_f(\|\mathbf{T}_k\|^2))^r \|\mathbf{T}_k\|^{2s}], \tag{2.47}$$

com $r, s = 1, 2$, e $s < r$.

Note que a estrutura que tem a matriz de informação dificulta a inferência dos parâmetros do modelo como, por exemplo, a determinação da variância assintótica do estimador de máxima verossimilhança do coeficiente angular $\beta$. Contudo, tal dificuldade pode ser resolvida aplicando-se o procedimento de Cox e Reid (1987), para obter-se uma nova parametrização do modelo, na qual o parâmetro de interesse $\beta$ e ortogonal aos parâmetros restantes. Consideremos a seguir a parametrização ortogonal estudada em Wong (1989) quando $\lambda_e$ é conhecido, e em Bolfarine e Cordani (1993) quando $\lambda_x$ é conhecida sob o modelo normal. Temos então

$$\phi_1 = \alpha + \beta\mu_x, \quad \phi_2 = \mu_x, \quad \phi_4 = \sigma_{uu},$$

$$\phi_3 = \begin{cases} \lambda_x\beta^2\sigma_{uu} + (\lambda_x + 1)\sigma_{ee} & \text{, se } \lambda_x \text{ é conhecido,} \\ (\beta^2 + \lambda_e)\sigma_{xx} + \lambda_e\sigma_{uu} & \text{, se } \lambda_e \text{ é conhecido.} \end{cases} \tag{2.48}$$

Para esta parametrização, sob o modelo elíptico considerado, temos que $\boldsymbol{\mu}(\boldsymbol{\phi}) = \boldsymbol{\mu}(\boldsymbol{\phi}_L)$ e

$$\mathbf{V} = V(\boldsymbol{\phi}_E) = \begin{cases} (\lambda_x + 1)^{-1} \begin{pmatrix} \phi_3 + (\lambda_x\beta)^2\phi_4 & (\lambda_x + 1)\lambda_x\beta\phi_4 \\ (\lambda_x + 1)\lambda_x\beta\phi_4 & (\lambda_x + 1)^2\phi_4 \end{pmatrix} & \text{, se } \lambda_x \text{ é conhecido,} \\ \\ (\beta^2 + \lambda_e)^{-1} \begin{pmatrix} \beta^2\phi_3 + \lambda_e^2\phi_4 & \beta(\phi_3 - \lambda_e\phi_4) \\ \beta(\phi_3 - \lambda_e\phi_4) & \phi_3 + \beta^2\phi_4 \end{pmatrix} & \text{, se } \lambda_e \text{ é conhecido,} \end{cases} \tag{2.49}$$

onde $\boldsymbol{\phi}_L = (\phi_1, \phi_2)^\top$ e $\boldsymbol{\phi}_E = (\phi_3, \phi_4, \beta)^\top$, com $\boldsymbol{\phi} = (\phi_1, \phi_2, \phi_3, \phi_4, \beta)^\top$. Em ambos os casos, $|\mathbf{V}| = \phi_3\phi_4$ e $\boldsymbol{\mu}(\boldsymbol{\phi})$ e $\mathbf{V}(\boldsymbol{\phi})$ são globalmente ortogonais, como acontece no modelo normal (veja Arellano-Valle, 1994). Assim, sob a parametrização ortogonal definida em (2.48), e de (2.46), com $\theta_j$ substituído por $\phi_j$, temos que a matriz de informação $\mathbf{K} = \mathbf{K}(\boldsymbol{\phi})$ é dada por

$$\mathbf{K} = \mathrm{Diag}(\mathbf{K}_L, \mathbf{K}_E),$$

onde $\mathbf{K}_L$ e $\mathbf{K}_E$ são as submatrizes de informação correspondentes aos parâmetros $\boldsymbol{\phi}_L$ e $\boldsymbol{\phi}_E$, respectivamente, que são dadas por

$$\mathbf{K}_L = \frac{4n}{p} a_f(2,1)\mathbf{V}^{-1} \quad \text{e} \quad \mathbf{K}_E = (K_{l,m}), \tag{2.50}$$

onde

$$K_{l,m} = \begin{cases} \frac{p}{2}\left[\frac{p+4\delta_{lm}}{p^2(p+2)} a_f(2,2) - \frac{1}{4}\right]\left(\frac{\phi_l\phi_m}{n}\right)^{-1} & , l, m = 3, 4, \\ \frac{4n}{p(p+2)} a_f(2,2)(\sigma_\beta^2)^{-1} & , l, m = \beta, \\ 0 & , l, m = 3, 4,\ m = 5, \text{ ou } l = 5,\ m = 3, 4, \end{cases}$$

com

$$\sigma_\beta^2 = \sigma_\beta^2(\boldsymbol{\phi}_E) = \begin{cases} \frac{\phi_3}{\lambda_x^2 \phi_4} & \text{, se } \lambda_x \text{ é conhecido,} \\ \left(\frac{\beta^2+\lambda_e}{\phi_3-\lambda_e\phi_4}\right)^2 \phi_3\phi_4 & \text{, se } \lambda_e \text{ é conhecido.} \end{cases} \tag{2.51}$$

A notação $\text{Diag}(\mathbf{A}_1, \ldots, \mathbf{A}_p)$ denota a matriz diagonal cujos elementos $\mathbf{A}_1, \ldots, \mathbf{A}_p$ são matrizes. Note que $\sigma_\beta^2$ pode ser escrito de maneira unificada como

$$\sigma_\beta^2 = \beta^2 \left(\frac{1-\rho_{YX}^2}{\rho_{YX}^2}\right), \tag{2.52}$$

onde $\rho_{YX} = \text{Cov}[Y_k, X_k]/(\text{Var}[Y_k]\text{Var}[X_k])^{1/2}$.

### 2.5.2 Estimação de máxima verossimilhança

Sob a condição $\lambda_x$ conhecido, consideremos o estimador obtido pelo método dos momentos,

$$\widehat{\beta}_N = \left(\frac{\lambda_x+1}{\lambda_x}\right)\frac{S_{YX}}{S_{XX}}. \tag{2.53}$$

Bolfarine e Cordani (1993) mostram que sob normalidade, $\widehat{\beta}_N$ é o estimador de máxima verossimilhança de $\beta$. O resultado seguinte mostra que o estimador $\widehat{\beta}_N$, sob o modelo elíptico independente é consistente e assintoticamente normal.

**Teorema 2.5** *Sob o modelo elíptico independente satisfazendo as condições S1 e S3 com $\lambda_x$ conhecido,*

$$\widehat{\beta}_N \xrightarrow{\text{q.c.}} \beta \quad e \quad \sqrt{n}(\widehat{\beta}_N - \beta) \xrightarrow{\text{d}} N(0, \Delta_N), \tag{2.54}$$

*onde*

$$\begin{aligned} \Delta_N &= \left(\frac{\lambda_x+1}{\lambda_x}\right)^2 (\kappa+1)\frac{(\sigma_{YY}\sigma_{XX}-\sigma_{XY}^2)}{\sigma_{XX}^2} \\ &= (\kappa+1)\beta^2\left(\frac{1-\rho_{XY}^2}{\rho^2 XY}\right). \end{aligned} \tag{2.55}$$

**Prova:** Note que $\widehat{\beta}_N = g(\text{Vech}(\mathbf{S}_n))$, onde $g(y,z,w) = \left(\frac{\lambda_x+1}{\lambda_x}\right)\frac{z}{w}$, $w \neq 0$. Por i) do Teorema 2.2, e da continuidade de $g$ temos que $\widehat{\beta}_N \xrightarrow{\text{q.c.}} \beta$. Por outro lado, de ii) do Teorema 2.2, sob o modelo estrutural, temos que

$$\sqrt{n}\mathrm{Vech}(\mathbf{S}_n - \boldsymbol{\Sigma}) \xrightarrow{\mathrm{d}} N(0, \boldsymbol{\nabla}),$$

onde $\boldsymbol{\nabla} = \mathbf{H}^\top \boldsymbol{\Lambda} \mathbf{H}$. Assim pelos resultado acima e usando o método delta, temos que

$$\sqrt{n}(\widehat{\beta}_N - \beta) \xrightarrow{\mathrm{d}} N(0, \mathbf{d}^\top \boldsymbol{\nabla} \mathbf{d}),$$

onde

$$\mathbf{d} = \frac{1}{\sigma_{XX}} \left(0, \frac{\lambda_x + 1}{\lambda_x}, -\beta\right)^\top$$

e

$$\mathbf{d}^\top \boldsymbol{\nabla} \mathbf{d} = \left(\frac{\lambda_x + 1}{\lambda_x}\right)^2 \left(\frac{\sigma_{YY}\sigma_{XX} - \sigma_{XY}^2}{\sigma_{XX}^2}\right).$$

A outra parte da igualdade segue facilmente da definição de $\rho_{XY}$. □

No teorema seguinte apresentamos em forma unificada a distribuição assintótica do estimador de máxima verossimilhança $\widehat{\beta}_f$ de $\beta$ sob a condição de $\lambda_e$ ou $\lambda_x$ conhecido.

**Teorema 2.6** *Sob o modelo estrutural independente com $\lambda_e$ ou $\lambda_x$ conhecido, a distribuição assintótica do estimador de máxima verossimilhança $\widehat{\beta}_f$ de $\beta$ é dado por*

$$\sqrt{n}(\widehat{\beta}_f - \beta) \to N(0, \sigma_\beta^2 / a_f), \tag{2.56}$$

*onde $\sigma_\beta^2$ é como em (2.52) e $a_f = \frac{4}{p(p+2)} a_f(2,2)$, com $a_f(2,2)$ obtido a partir de (2.47).*

**Prova:** A demonstração segue dos resultados considerados anteriormente com relação à matriz de informação de Fisher. □

Para o modelo estrutural normal, temos que $a_f = 1$ e

$$\beta_f = \begin{cases} \widehat{\beta}_N & \text{, se } \lambda_x \text{ é conhecido,} \\ \widehat{\beta}_{GLS} & \text{, se } \lambda_e \text{ é conhecido,} \end{cases}$$

onde $\widehat{\beta}_{GLS}$ e $\widehat{\beta}_N$ são como em (2.33) e (2.53), respectivamente.

**Corolário 2.2** *Sob o modelo estrutural elíptico independente, com $\lambda_e$ ou $\lambda_x$ conhecido, a eficiência relativa assintótica do estimador de máxima verossimilhança com respeito ao estimador $\widehat{\beta}_{GLS}$ ou $\widehat{\beta}_N$ é dado por*

$$e_{\widehat{\beta}_{GLS},\widehat{\beta}_f} = e_{\widehat{\beta}_N,\widehat{\beta}_f} = \frac{1}{(\kappa+1)a_f}. \tag{2.57}$$

**Prova:** A demonstração segue do Teorema 2.3 sob o modelo estrutural (veja seção 2.4), do Teorma 2.5, juntamente com o Teorema 2.6 e das relações

$$e_{\widehat{\beta}_{GLS},\widehat{\beta}_f} = \frac{\Delta_f}{\Delta_{GLS}}, \text{ se } \lambda_e \text{ é conhecido}$$

e

$$e_{\widehat{\beta}_N,\widehat{\beta}_f} = \frac{\Delta_f}{\Delta_N}, \text{ se } \lambda_x \text{ é conhecido.}$$

□

Para o modelo normal, $W_f(u) = -\frac{1}{2}$, $\kappa = 0$ e $a_f = 1$. Assim, $\Delta_f = \Delta_{GLS}$, se $\lambda_e$ conhecido e $\Delta_f = \Delta_N$, se $\lambda_x$ conhecido. Para o modelo $t$-Student, isto é, $\mathbf{Z} \sim t_2(\boldsymbol{\mu}, \mathbf{V}; \nu)$, onde $\boldsymbol{\mu} = E[\mathbf{Z}]$ e $\boldsymbol{\Sigma} = \mathrm{Var}[\mathbf{Z}] = (\nu/\nu-2)\mathbf{V}$, $\nu > 2$, temos que

$$W_f(u) = -\frac{\nu+2}{2}(\nu+u)^{-1},\ u>0,\ a_f = \frac{\nu+2}{\nu+4} \quad \text{e} \quad \kappa+1 = \frac{\nu-2}{\nu-4},\ \nu>4$$

e neste caso de (2.50), a matriz de informação sob a parametrização ortogonal é dada por $\mathbf{K} = \mathrm{Diag}(\mathbf{K}_L, \mathbf{K}_E)$ (veja Arellano-Valle, 1994), com

$$\mathbf{K}_L = \left(\frac{\nu+2}{\nu+4}\right)\left(\frac{1}{n}\mathbf{V}\right)^{-1}$$

e

$$\mathbf{K}_E = \begin{pmatrix} K_{3,3} & K_{3,4} & 0 \\ K_{3,4} & K_{4,4} & 0 \\ 0 & 0 & K_{\beta,\beta} \end{pmatrix} = \begin{pmatrix} \frac{\nu+2}{\nu+4}\left(\frac{2\phi_3^2}{n}\right)^{-1} & -\frac{1}{\nu+4}\left(\frac{2\phi_3\phi_4}{n}\right)^{-1} & 0 \\ -\frac{1}{\nu+4}\left(\frac{2\phi_3\phi_4}{n}\right)^{-1} & \frac{\nu+2}{\nu+4}\left(\frac{2\phi_4^2}{n}\right)^{-1} & 0 \\ 0 & 0 & \frac{\nu+2}{\nu+4}\left(\frac{\sigma_\beta^2}{n}\right)^{-1} \end{pmatrix},$$

onde $\sigma_\beta^2$ é como em (2.52). Assim, a eficiência assintótica do estimador de máxima verossimilhança $\widehat{\beta}_f$ com respeito aos estimadores $\widehat{\beta}_{GLS}$ e $\widehat{\beta}_N$ é dada por

$$e_{\widehat{\beta}_{GLS},\widehat{\beta}_f} = e_{\widehat{\beta}_N,\widehat{\beta}_f} = \frac{(\nu+4)(\nu-4)}{(\nu+2)(\nu-2)} < 1,\ \nu > 4,$$

isto é, o estimador de máxima verossimilhança $\widehat{\beta}_f$ é assintoticamente mais eficiente que os estimadores $\widehat{\beta}_{GLS}$ e $\widehat{\beta}_f$ sob as condições $\lambda_e$ e $\lambda_x$, respectivamente. Note que sob normalidade $e_{\widehat{\beta}_{GLS},\widehat{\beta}_f} = e_{\widehat{\beta}_N,\widehat{\beta}_f} = 1$, ou seja, $\widehat{\beta}_f = \widehat{\beta}_{GLS}$ se $\lambda_e$ conhecido e $\widehat{\beta}_f = \widehat{\beta}_N$, se $\lambda_x$ é conhecido.

**Teorema 2.7** *Sob o modelo estrutural independente com $\sigma_{uu}$ conhecido, a distribuição assintótica do estimador de máxima verossimilhança $\widehat{\beta}_f$ de $\beta$ é dada por*

$$\sqrt{n}(\widehat{\beta}_f - \beta) \xrightarrow{\mathrm{d}} N(0, \Delta_f), \tag{2.58}$$

*onde*

$$\Delta_f = \frac{1}{a_f}\left[\frac{\sigma_{XX}\sigma_{YY} - \sigma_{XY}^2}{\sigma_{xx}^2} + \frac{3a_f - 1}{2a_f - 1}\frac{(\beta\sigma_{XX} - \sigma_{XY})^2}{\sigma_{xx}^2}\right], \tag{2.59}$$

*com $a_f$ como no Teorema 2.6.*

**Prova:** Considerando a parametrização

$$\phi_1 = \mu_x, \ \phi_3 = \alpha + \beta\mu_x, \ \phi_3 = \beta, \ \phi_4 = \sigma_{ee}, \ \phi_5 = \sigma_{xx},$$

temos, de (2.46) que a matriz de informação tem a forma

$$\mathbf{K} = K(\boldsymbol{\phi}) = \mathrm{Diag}(\mathbf{K}_L, \mathbf{K}_S),$$

onde $\mathbf{K}_L = (K_{l,m})$, $l, m = 1, 2$; $\mathbf{K}_S = (K_{l,m})$, $l, m = 3, 4, 5$. Assim, a variância assintótica do estimador $\sqrt{n}\widehat{\beta}_M$ é dada pelo elemento $K_{11}^S$ correspondente à inversa da submatriz de informação $\mathbf{K}_S$. Após algumas manipulações algébricas, temos então que

$$\Delta_f = \frac{1}{a_f}\left[\frac{\sigma_{XX}\sigma_{YY} - \sigma_{XY}^2}{\sigma_{xx}^2} + \frac{3a_f - 1}{2a_f - 1}\frac{(\beta\sigma_{XX} - \sigma_{XY})^2}{\sigma_{xx}^2}\right],$$

onde $a_f = \frac{4}{p(p+2)}a_f(2, 2)$.

**Corolário 2.3** *Sob o modelo estrutural independente satisfazendo as condições S1 e S3 e com $\sigma_{uu}$ conhecido, a eficiência relativa assintótica do estimador de máxima verossimilhança com respeito a $\widehat{\beta}_M$ é dada por*

$$e_{\widehat{\beta}_M,\widehat{\beta}_f} = \frac{1}{a_f(\kappa+1)}\left\{\frac{1+\frac{3a_f+1}{2a_f-1}a_{YX}}{1+\frac{3x+2}{k+1}a_{YX}}\right\}, \tag{2.60}$$

*onde*

$$a_{YX} = \frac{\rho_{YX}^2}{\lambda_x^2(1-\rho_{YX}^2)}.$$

**Prova:** A prova segue do Teorema 2.4 sob o modelo estrutural, do teorema anterior e a relação

$$e_{\widehat{\beta}_M,\widehat{\beta}_f} = \frac{\Delta_f}{\Delta_M},$$

onde $\Delta_f$ e $\Delta_M$ são as variâncias assintóticas do estimador de máxima verossilhança $\widehat{\beta}_f$ e do estimador $\widehat{\beta}_M$, respectivamente.

Note que sob normalidade, $a_f = 1$ e $\kappa = 0$, de modo que $e_{\widehat{\beta}_M,\widehat{\beta}_f} = 1$. Neste caso, $\widehat{\beta}_M = \widehat{\beta}_f$, isto é, o estimador de máxima verossimilhança $\widehat{\beta}_f$ é equivalente ao estimador $\widehat{\beta}_M$. Agora, sob o modelo $t$-Student $\kappa+1 = (\nu-2)/(\nu-4)$ e $a_f = (\nu+2)/(\nu+4)$. Portanto,

$$e_{\widehat{\beta}_M,\widehat{\beta}_f} = \frac{(\nu+4)(\nu-4)}{(\nu+2)(\nu-2)}\left\{\frac{1+2\left(\frac{\nu+1}{\nu}\right)a_{YX}}{1+2\left(\frac{\nu-1}{\nu-2}\right)a_{YX}}\right\}, \quad \nu > 4,$$

o que mostra que o estimador de máxima verossimilhança $\widehat{\beta}_f$ é assintoticamente mais eficiente que o estimador $\widehat{\beta}_M$.

## 2.6 O modelo estrutural elíptico dependente

Os modelos elípticos dependentes no contexto geral; são caracterizados ao assumir que a distribuição conjunta das variáveis de interesse é elíptica. Assim, sendo $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ uma seqüência de vetores aleatórios de dimensão $p$, o modelo elíptico dependente é definido supondo que

$$\mathbf{Z} = (\mathbf{Z}_1^\top, \ldots, \mathbf{Z}_n^\top)^\top \sim E\ell_{pn}(\boldsymbol{\mu}, \boldsymbol{\Psi}; \phi),$$

com função densidade (se existir) dada por

$$|\boldsymbol{\Psi}|^{-\frac{1}{2}} f((\mathbf{Z} - \boldsymbol{\mu})^\top \boldsymbol{\Psi}^{-1} (\mathbf{Z} - \boldsymbol{\mu})),$$

com $f(u) \geq 0$. Neste caso, pelo Teorema 1 dado em Anderson, Fang e Hsu (1986), o estimador de máxima verossimilhança de $\boldsymbol{\mu}$ e $\boldsymbol{\Psi}$ são dados por

$$\widehat{\boldsymbol{\mu}} = \widehat{\boldsymbol{\mu}}_N \quad \text{e} \quad \widehat{\boldsymbol{\Psi}} = \frac{np}{u_f} \widehat{\boldsymbol{\Psi}}_N, \; n > p, \tag{2.61}$$

onde $\widehat{\boldsymbol{\mu}}_N$ e $\widehat{\boldsymbol{\Psi}}_N$ são os respectivos estimadores de máxima verossimilhança sob normalidade e $\mu_f$ é o ponto de máximo da função $u^{np/2} f(u)$. Por exemplo, se $f(u) = k(p, \nu)\nu^{\nu/2}\{\nu + u\}^{-(\nu+p)/2}$ (o modelo $t$-Student com $\nu$ graus de liberdade) é fácil ver que $u_f = np$, para todo $\nu > 0$. Em geral, se $f(u)$, $u \geq 0$, é uma função contínua e decrescente para $u$ suficientemente grande, então a função $u^{np/2} f(u)$ tem ponto de máximo $u_f$. Este fato vale em particular quando a densidade esférica $f(\mathbf{z}^\top \mathbf{z})$, $\mathbf{z} \in \mathbb{R}^{np}$, tem segundos momentos finitos (veja Anderson et al., 1986).

No contexto da teoria de inferência baseada no contexto elíptico, o resultado de Anderson et al. (1986) tem grande importância, e será de grande utilidade nesta seção. De fato, se $\mathbf{X} \sim E\ell_p(\mathbf{0}, \mathbf{I}_p; \phi)$, $P(\mathbf{X} = 0) = 0$ e $\delta(\mathbf{X})$ é uma estatística tal que $\delta(a\mathbf{X}) \stackrel{\mathrm{d}}{=} \delta(\mathbf{X})$, para todo $a > 0$, então (veja Fang et al., 1990)

$$\delta(\mathbf{X}) \stackrel{\mathrm{d}}{=} \delta(\mathbf{Z}), \quad \mathbf{Z} \sim N_p(\mathbf{0}, \mathbf{I}_p). \tag{2.62}$$

O modelo elíptico estrutural dependente com erro nas variáveis é obtido supondo que a seqüência $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$, definida em (2.25) com $\boldsymbol{\eta}_k = \boldsymbol{\eta}$ ($\mu_{x_k} = \mu_x$, $k = 1, \ldots, n$ e $\lambda_x^* = 0$) é tal que

$$\mathbf{Z} = (\mathbf{Z}_1^\top, \ldots, \mathbf{Z}_n^\top)^\top \sim E\ell_{2n}(\mathbf{1}_n \otimes \boldsymbol{\mu}, \mathbf{I}_n \otimes \mathbf{V}; \phi),$$

com densidade dada por

$$|\mathbf{V}|^{-n/2} f\left(\sum_{k=1}^{n} (\mathbf{Z}_k - \boldsymbol{\mu})^\top \mathbf{V}^{-1} (\mathbf{Z}_k - \boldsymbol{\mu})\right),$$

onde $f(u) \geq 0$, $u > 0$ e

$$\boldsymbol{\mu} = \mathbf{a} + \mathbf{B}\boldsymbol{\eta} \quad \text{e} \quad \mathbf{V} = \mathbf{B}\boldsymbol{\Psi}\mathbf{B}^\top,$$

com $\mathbf{a}$, $\mathbf{B}$, $\boldsymbol{\Psi}$ e $\boldsymbol{\eta}$ como em (2.25).

Lembremos que, sob o modelo normal, a estatística $\widehat{\beta}_{GLS}$, $\widehat{\beta}_M$ ou $\widehat{\beta}_N$ (definida em (2.33), (2.37) e (2.53), respectivamente) é o estimador de máxima verossimilhança de $\beta$ sob a respectiva condição $\lambda_e$, $\lambda_x$ ou $\sigma_{uu}$ conhecido. No que segue, mostraremos que tais estimadores são também estimadores de máxima verossimilhança de $\beta$ sob o modelo elíptico dependente.

**Teorema 2.8** *Sob o modelo elíptico dependente, os estimadores $\widehat{\beta}_{GLS}$, $\widehat{\beta}_N$ e $\widehat{\beta}_M$ de $\beta$ são os estimadores de máxima verossimilhança sob as respectivas condições $\lambda_e$, $\lambda_x$ e $\sigma_{uu}$ conhecidos.*

**Prova:** Primeiramente note que os estimadores $\widehat{\beta}_{GLS}$, $\widehat{\beta}_N$ e $\widehat{\beta}_M$ satisfazem a condição de invariância considerada acima. Assim, por (2.62) (veja Fang et al., 1986), temos que

$$\widehat{\beta}_{GLS}(\mathbf{Z}) \stackrel{\mathrm{d}}{=} \widehat{\beta}_{GLS}(\mathbf{Z}_N), \; \widehat{\beta}_N(\mathbf{Z}) \stackrel{\mathrm{d}}{=} \widehat{\beta}_N(\mathbf{Z}_N) \quad \text{e} \quad \widehat{\beta}_M(\mathbf{Z}) \stackrel{\mathrm{d}}{=} \widehat{\beta}_M(\mathbf{Z}_N),$$

onde $\mathbf{Z}_N \sim N_{2n}(\mathbf{1}_n \otimes \boldsymbol{\mu}, \mathbf{I}_n \otimes \mathbf{V})$. Agora, sejam $\widehat{\beta}_f$ e $\widehat{\beta}$ os estimadores de máxima verossimilhança de $\beta$ sob o modelo elíptico dependente e normal, sob a condição $\lambda_e$, $\lambda_x$ ou $\sigma_{uu}$ conhecido, respectivamente. Então, pelo resultado em (2.61), segue que

$$\widehat{\beta}_f = \widehat{\beta} = \begin{cases} \widehat{\beta}_{GLS} & \text{, se } \lambda_e \text{ conhecido,} \\ \widehat{\beta}_N & \text{, se } \lambda_x \text{ conhecido,} \\ \widehat{\beta}_M & \text{, se } \sigma_{uu} \text{ conhecido,} \end{cases}$$

isto é, que sob as condições $\lambda_e$, $\lambda_x$ ou $\sigma_{uu}$ conhecido, os respectivos estimadores $\widehat{\beta}_{GLS}$, $\widehat{\beta}_N$ ou $\widehat{\beta}_M$ são os estimadores de máxima verossimilhança sob os correspondentes modelos elípticos dependentes.

# Capítulo 3

# O modelo funcional elíptico

## 3.1 Introdução

Neste capítulo estudaremos o modelo com erros nas variáveis definido por

$$\mathbf{Y}_i = \boldsymbol{\alpha} + \boldsymbol{\beta} x_i + \mathbf{e}_i, \quad i = 1, \ldots, n, \tag{3.1}$$

onde $x_1, \ldots, x_n$ são quantidades não aleatórias medidas com erro, isto é, assumindo que em lugar do verdadeiro valor $x_i$, observamos $X_i$, tal que

$$X_i = x_i + u_i, \quad i = 1, \ldots, n, \tag{3.2}$$

onde $u_i$ representa o erro de medição da quantidade $x_i$. Neste capítulo vamos supor que $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_q)^\top$ e $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_q)^\top$ são vetores de parâmetros $q$-dimensional, $q \geq 1$ e $X_i$ uma variável aleatória real.

O modelo definido em (3.1) e (3.2) pode ser expresso em forma matricial como

$$\mathbf{Z}_i = \mathbf{a} + \mathbf{b} x_i + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, n, \tag{3.3}$$

onde $\mathbf{Z}_i = (\mathbf{Y}_i^\top, X_i)^\top$, $\boldsymbol{\epsilon}_i = (\mathbf{e}_i^\top, u_i)^\top$, $\mathbf{a} = (\boldsymbol{\alpha}^\top, 0)^\top$ e $\mathbf{b} = (\boldsymbol{\beta}^\top, 1)^\top$ são vetores de dimensão $p = q + 1$.

Vamos supor que a seqüência $\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n$ de vetores aleatórios $p$-dimensional são independentes e identicamente distribuídos com

$$\boldsymbol{\epsilon}_j \sim E\ell_p(\mathbf{0}, \mathbf{I}_p; f), \quad p = q + 1, \tag{3.4}$$

onde $f(u)$, $u \geq 0$ é uma função não negativa tal que $r^{p-1}f(r^2)$ é integrável sobre $[0, \infty)$, isto é, em (3.4) estamos supondo que os erros $\epsilon_1, \ldots, \epsilon_n$ têm densidade $p$-variada, $f(\mathbf{x}^\top \mathbf{x})$, $\mathbf{x} \in \mathbb{R}^p$. Assim, pela propriedade das distribuições elípticas dada no Teorema A.3 do Apêndice A, temos de (3.3) que

$$\mathbf{Z}_i \sim E\ell_p(\mathbf{a} + \mathbf{b}x_i, \mathbf{I}_p; f), \tag{3.5}$$

com função densidade dada por

$$f_i(\mathbf{Z}_i; \boldsymbol{\theta}, x_i) = f(\|\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_i\|^2),$$

onde $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top)^\top$ é o vetor de parâmetros estruturais do modelo.

O modelo definido em (3.5) será chamado de "Modelo Linear Funcional Elíptico com Erros nas Variáveis" ou simplesmente Modelo Funcional Elíptico. Sob normalidade, o modelo em (3.3) é chamado "Modelo Linear Funcional com Erros nas Variáveis" ou simplesmente Modelo Funcional (Kendall e Stuart, 1970, e Fuller, 1987) e passa a ser um caso particular do modelo funcional elíptico, desde que a distribuição normal está contida na classe das distribuições elípticas.

Como é habitual, o principal objetivo é fazer inferências sobre os parâmetros estruturais $\boldsymbol{\alpha}$ e $\boldsymbol{\beta}$ a partir das observações $(\mathbf{Y}_i^\top, X_i)$, $i = 1, \ldots, n$. A diferença com o modelo estrutural definido na Seção 2.5 do Capítulo 2 é que no modelo funcional os $x_i$'s são parâmetros fixos e portanto o número de parâmetros cresce com o tamanho da amostra, $n$. Para $q = 1$ e supondo que $\epsilon_i \overset{\text{iid}}{\sim} N_2\left(\mathbf{0}, \begin{pmatrix} \sigma_{ee} & 0 \\ 0 & \sigma_{uu} \end{pmatrix}\right)$, $i = 1, \ldots, n$, o logaritmo da função de verossimilhança do modelo funcional correspondente a uma amostra de tamanho $n$, tem como característica principal ser ilimitado, de modo que não existe solução de máxima verossimilhança. Solari (1969) mostrou que a raiz da equação da primeira derivada com respeito ao parâmetro estrutural é um ponto de sela e não de um máximo. Zellner (1971) estudou o problema da não existência do máximo do logaritmo da função de verossimilhança no espaço paramétrico, mostrando que a solução para $\beta$ não pertence à região admissível do espaço paramétrico, a qual é dada por

$$\frac{|\mathrm{Cov}[X,Y]|}{\sigma_{XX}} \leq |\beta| \leq \frac{\sigma_{YY}}{|\mathrm{Cov}[X,Y]|},$$

com $\mathrm{Cov}[X,Y] \neq 0$. Suposições adicionais são, portanto, requeridas para tornar viável o estudo de inferências no modelo funcional normal. Algumas suposições comumente adotadas são:

i) $\sigma_{ee}$ e $\sigma_{uu}$ conhecidos;

ii) $\sigma_{ee}$ ou $\sigma_{uu}$ conhecido;

iii) $\lambda_e = \sigma_{ee}/\sigma_{uu}$ conhecido.

A condição iii) é mais usada na literatura, como é considerado em Lindley e El Sayad (1968), Zellner (1971), Kendall e Stuart (1970), Mak (1982), Glesser (1981), dentre outros. A condição ii) é usada, por exemplo, em Cheng e Van Ness (1991). Sob a suposição i), Barnett (1969) mostra que os estimadores de máxima verossimilhança (EMV) de $\alpha$, $\beta$, $x_1, \ldots, x_n$, são dados por

$$\begin{aligned}
\widehat{\alpha} &= \overline{Y} - \widehat{\beta}\overline{X}, \\
\widehat{x}_i &= \frac{\widehat{\beta}(Y_i - \widehat{\alpha}) + \lambda_e X_i}{\widehat{\beta}^2 + \lambda_e}, \quad i = 1, \ldots, n, \\
\widehat{\beta} &= \frac{S_{YY} - \lambda_e S_{XX} + \sqrt{(S_{YY} - \lambda_e S_{XX})^2 + 4\lambda_e S_{XY}^2}}{2S_{XY}},
\end{aligned}$$

onde $S_{XX}$, $S_{YY}$ e $S_{XY}$ são as entradas da matriz de covariâncias amostrais $\mathbf{S}_n$ e $\widehat{\beta}$ depende de $\sigma_{ee}$ e $\sigma_{uu}$ somente através de $\phi$. Sob a suposição iii), Glesser (1981) mostra que o EMV de $\beta$ é dado como em (2.33) (veja o capítulo 2). Em situações onde o vetor de erros tem distribuição elíptica não normal (por exemplo, $t$-Student), não é possível obter uma forma explícita para o EMV. Em tais situações, um método comumente usado para obter tais estimadores é dado pelo algoritmo EM (ver Arellano-Valle, 1994). Mesmo sob normalidade, a distribuição exata dos estimadores é bastante dificultada devido a presença dos parâmetros incidentais $x_1, \ldots, x_n$ como é apontado, dentre outros, por Neyman e Scott (1948, 1951), Kiefer et al. (1956), Zellner (1971), Gleser (1981) e Fuller (1987). Embora o tratamento de Gleser (1981) e Fuller (1987) seja bastante geral, ele está baseado na teoria de vetores e valores próprios e pode apresentar algumas dificuldades se o leitor estiver interessado em aplicações.

No modelo funcional sob normalidade com $q = 1$, Patefield (1976) mostra que a matriz de covariâncias assintótica do EMV não corresponde à inversa da matriz de informação de Fisher. Contudo, Mak (1982) sob um um modelo bem geral, estabeleceu certas condições gerais para garantir a consistência e distribuição assintótica do EMV do parâmetro estrutural. Usando estes resultados, Mak (1982) estudou as propriedades assintóticas dos EMV dos parâmetros estruturais no modelo definido em (3.3) para $q = 1$ e supondo que $\epsilon_1 \sim N_2(\mathbf{0}, \sigma^2\mathbf{I}_2)$, onde $\sigma^2$ é considerado desconhecido, o que equivale a supor a condição de identificabilidade iii), isto é, a supor que $\lambda_e = \sigma_{ee}/\sigma_{uu}$ é conhecido. No caso multivariado, um estudo é feito em Mak (1982) supondo $\alpha = 0$ e $\boldsymbol{\epsilon}_1 \sim N_p(\mathbf{0}, \sigma^2\mathbf{I}_p)$, sendo $\sigma^2$ considerado conhecido.

O modelo definido em (3.3) com $q \geq 1$ é usado no estudo de calibração comparativa funcional (veja Williams, 1969, Kimura, 1992), que trata da comparação de instrumentos ou métodos de medição, quando cada um deles é usado para medir uma característica num grupo comum de unidades experimentais. A necessidade de comparar instrumentos de medição que diferem em custos, velocidades, ou outros fatores, tem aparecido com freqüência em diversas áreas tais como em medicina, psicologia, educação e agricultura, entre outros. Por exemplo, Grubbs (1948, 1973) compara a eficiência de três cronômetros. Na área de medicina, citamos, entre outros, os trabalhos de Barnett (1969), Kelly (1984, 1985) e Kaak et al. (1994). Fuller (1987) apresenta exemplos na área da agricultura e Dunn (1992) considera aplicações em psicologia e educação. Para efeito de aplicações, Kimura (1992) propõe um algoritmo tipo EM para estimar os parâmetros $\alpha$ e $\beta$. Para estimar a matriz de covariância assintótica, Kimura (1992) propõe a inversa da matriz de informação observada. Bolfarine e Galea-Rojas (1995) mostram que a matriz de covariâncias assintóticas não coincide com a inversa da matriz de informação de Fisher.

Neste capítulo trataremos o modelo de calibração comparativa no caso funcional. Especificamente, estudaremos a consistência e normalidade assintótica dos estimadores de máxima verossimilhança dos parâmetros estruturais do modelo estatístico definido em (3.5).

## 3.2 Notações e resultados preliminares

Nesta seção apresentamos de forma resumida os principais resultados de Mak (1982), que serão utilizados no estudo das propriedades assintóticas dos EMV dos parâmetros estruturais do modelo funcional elíptico definido em (3.5).

Seja $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ uma seqüência de vetores aleatórios independentes $p$-dimensionais, onde para cada $i = 1, \ldots, n$, $\mathbf{Z}_i$ tem função densidade $f_i(\mathbf{z}_i; \boldsymbol{\theta}, x_i)$, com $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^\top \subset \boldsymbol{\Theta} \subseteq \mathbb{R}^p$ e $x_i \in \chi_i \subset \mathbb{R}$, $i = 1, \ldots, n$. Suponhamos também que $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ e $x_{i0} \in \chi_i$, $i = 1, \ldots, n$, onde $\boldsymbol{\theta}_0$ e $x_{10}, \ldots, x_{n0}$ denotam os verdadeiros valores de $\boldsymbol{\theta}$ (parâmetro estrutural) e $x_1, \ldots, x_n$ (parâmetros incidentais), respectivamente. Assim, para uma amostra de tamanho $n$, o logaritmo da função de verossimilhança é dada por

$$\ell(\boldsymbol{\theta}, \mathbf{x}) = \sum_{i=1}^{n} \log f_i(\mathbf{z}_i; \boldsymbol{\theta}, x_i), \tag{3.6}$$

onde $\mathbf{x} = (x_1, \ldots, x_n)^\top$, com $x_i \in \chi_i$, $i = 1, \ldots, n$, e $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

Suponha que para cada $i$, dado $\boldsymbol{\theta}$, existe um estimador $\widetilde{x}_i = \widetilde{x}_i(\mathbf{Z}_i; \boldsymbol{\theta})$ (dependendo de $\boldsymbol{\theta}$) para $x_i$. Em particular, $\widetilde{x}_i$ pode ser o EMV condicional de $x_i$ dado $\boldsymbol{\theta}$. Substituindo $x_i$

por $\widetilde{x}_i$ em (3.6), temos que

$$\ell(\boldsymbol{\theta}, \widetilde{x}) = \sum_{i=1}^{n} \log f_i(\mathbf{z}_i; \boldsymbol{\theta}, \widetilde{x}_i) = \sum_{i=1}^{n} h_i(\mathbf{z}_i; \boldsymbol{\theta}), \tag{3.7}$$

onde $h_i(\mathbf{z}_i; \theta) = \log f_i(\mathbf{z}_i; \theta, \widetilde{x}_i)$, $i = 1, \ldots, n$.

No resto do capítulo, as esperanças serão tomadas sob os verdadeiros valores $\boldsymbol{\theta}_0$ e $x_{10}, \ldots, x_{n0}$, e serão denotadas por

$$E_0[\cdot] = E[\cdot|\boldsymbol{\theta}_0, x_{10}, \ldots, x_{n0}]. \tag{3.8}$$

Supondo que as derivadas envolvidas existam quase certamente para todo $i$, definimos

$$\begin{aligned}
q_{i\theta_j}(\mathbf{z}_i; \boldsymbol{\theta}) &= \frac{\partial}{\partial \theta_j} h(\mathbf{z}_i; \boldsymbol{\theta}), \quad j = 1, \ldots, p, & (3.9)\\
q_{i\theta_k\theta_j}(\mathbf{z}_i; \boldsymbol{\theta}) &= \frac{\partial^2}{\partial \theta_j \theta_k} h(\mathbf{z}_i; \boldsymbol{\theta}), \quad j, k = 1, \ldots, p, & (3.10)\\
q_{i\theta_k\theta_j}(\mathbf{z}_i; \boldsymbol{\theta}) &= (q_{i\theta_k}(\mathbf{z}_i; \boldsymbol{\theta}))(q_{i\theta_j}(\mathbf{z}_i; \boldsymbol{\theta})), \quad k, j = 1, \ldots, p. & (3.11)
\end{aligned}$$

Além disso, sejam $E_0[A_n(\boldsymbol{\theta})] = (a_{kj}^n(\boldsymbol{\theta}))$ e $V_n(\boldsymbol{\theta}) = (\nu_{kj}^n(\boldsymbol{\theta}))$ as matrizes simétricas de dimensão $p \times p$ definidas por

$$a_{kj}^n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n} E_0[q_{i\theta_k\theta_j}(\mathbf{Z}_i; \boldsymbol{\theta})] \tag{3.12}$$

$$\begin{aligned}
\nu_{kj}^n(\boldsymbol{\theta}) &= \frac{1}{n} \sum_{i=1}^{n} \mathrm{Cov}_0[q_{i\theta_k}(\mathbf{Z}_i; \boldsymbol{\theta}), q_{i\theta_j}(\mathbf{Z}_i; \boldsymbol{\theta})]\\
&= \frac{1}{n} \sum_{i=1}^{n} \mathrm{Cov}_0\left[\frac{\partial h_i(\mathbf{Z}_i; \boldsymbol{\theta})}{\partial \theta_k}, \frac{\partial h_i(\mathbf{Z}_i; \boldsymbol{\theta})}{\partial \theta_j}\right], & (3.13)
\end{aligned}$$

respectivamente.

Suponhamos que exista $\boldsymbol{\theta}_1 \in \Theta^0$ que maximize a função

$$\overline{\psi}(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n} E_0[h_i(\mathbf{Z}_i; \boldsymbol{\theta})], \tag{3.14}$$

para $n$ suficientemente grande. Mak (1982) estabeleceu certas condições gerais sob as quais (3.7) tem um máximo $\widetilde{\boldsymbol{\theta}}_n$ que converge em probabilidade para $\boldsymbol{\theta}_1$ e tem distribuição assintótica tal que

$$\sqrt{n}(V_n(\boldsymbol{\theta}_1))^{-1/2}(E_0[A_n(\boldsymbol{\theta}_1)])(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_1) \xrightarrow{\mathrm{d}} N_p(\mathbf{0}, \mathbf{I}_p). \qquad (3.15)$$

Mak (1982) notou também que em algumas situações é possível obter estimadores $\widetilde{x}_i$ tais que $\boldsymbol{\theta}_1$ depende somente de $\boldsymbol{\theta}_0$ e é independente de $x_{i0}$, isto é, $\boldsymbol{\theta}_1 = g(\boldsymbol{\theta}_0)$, para alguma função $g$. Se $g$ é uma função um a um, então um estimador consistente de $\boldsymbol{\theta}_0$ é dado por $\widehat{\boldsymbol{\theta}}_n = g^{-1}(\widetilde{\boldsymbol{\theta}}_n)$. No caso em que $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_0$, o estimador consistente de $\boldsymbol{\theta}_1$ é também um estimador consistente de $\boldsymbol{\theta}_0$.

## 3.3 O modelo funcional elíptico univariado

Nesta seção estudaremos o comportamento assintótico do estimador de máxima verossimilhança do parâmetro estrutural do modelo funcional definido em (3.3), com $q = 1$, isto é, em (3.3) tem-se que

$$\mathbf{Z}_i = \mathbf{a} + \mathbf{b}x_i + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, n, \qquad (3.16)$$

onde $\mathbf{Z}_i = (Y_i, X_i)^\top$, $\boldsymbol{\epsilon}_i = (e_1, u_i)^\top$, $\mathbf{a} = (\alpha, 0)^\top$ e $\mathbf{b} = (\beta, 1)^\top$ são vetores bidimensionais de modo que em (3.4) estamos supondo que a seqüência de vetores de erros $\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n$ são independentes e têm uma distribuição bivariada comum $E\ell_2(\mathbf{0}, \mathbf{I}_2; f)$.

No que segue a seção será dividida em três subseções. Na Seção 3.3.1, estudaremos a consistência e a normalidade assintótica do EMV do parâmetro estrutural do modelo funcional. Na Seção 3.3.2, estudaremos a eficiência relativa assintótica do EMV, com respeito ao estimador de mínimos quadrados generalizados. Finalmente, na Seção 3.3.3, estudaremos com maiores detalhes o caso particular do modelo funcional $t$-Student, isto é, quando os erros do modelo considerado têm distribuição comum $t$-Student.

De (3.6) segue que para uma amostra de tamanho $n$, a função de log-verossimilhança é dada por

$$\ell(\boldsymbol{\theta}, \mathbf{x}) = \sum_{i=1}^{n} \log f(\|\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_i\|^2), \tag{3.17}$$

onde $\mathbf{x} = (x_1, \ldots, x_n)^\top$ e $\boldsymbol{\theta} = (\alpha, \beta)^\top$.

Para efeito de inferência, vamos supor que $f$ satisfaz as seguintes condições de regularidade:

A1. $f \in C^{(2)}$ e é decrescente em $(0, +\infty)$;

A2. $r^5 f(r^2)$ é integrável sobre $[0, +\infty)$, isto é, a distribuição de $\boldsymbol{\epsilon}_1$ admite quartos momentos finitos.

A seguinte notação será usada nesta seção. Para qualquer função $\Phi = \Phi(\boldsymbol{\theta})$ (que eventualmente pode ser uma matriz), denotaremos por $\Phi_0 = \Phi(\boldsymbol{\theta}_0)$, isto é, a função $\Phi$ avaliada no ponto $\boldsymbol{\theta}_0$. Sejam também

$$\mathbf{b}_1 = \frac{1}{\|\mathbf{b}\|}\begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \mathbf{b}_2 = \frac{1}{\|\mathbf{b}\|}\begin{pmatrix} 1 \\ -\beta \end{pmatrix}, \quad \mathbf{A} = \frac{1}{\|\mathbf{b}\|}\begin{pmatrix} 1 & -\beta \\ 0 & 1 \end{pmatrix} \tag{3.18}$$

e as variáveis aleatórias

$$s_i = \mathbf{b}_1^\top(\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_{i0}) \quad \text{e} \quad t_i = \mathbf{b}_2^\top(\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_{i0}), \tag{3.19}$$

$i = 1, \ldots, n$. As variáveis aleatórias $t_i$ e $s_i$ (que dependem do parâmetro $\boldsymbol{\theta}$) tem distribuição conjunta elíptica (não singular), como veremos a seguir.

**Lema 3.1** *Para cada $i$, temos que*

*i)* $\begin{pmatrix} t_i \\ s_i \end{pmatrix} \sim E\ell_2(\mathbf{A}[(\mathbf{a}_0 - \mathbf{a}) + (\mathbf{b}_0 - \mathbf{b})x_{i0}], \mathbf{A}\mathbf{A}^\top)$,
*onde* $\mathbf{A}\mathbf{A}^\top = \frac{1}{\|\mathbf{b}\|^2}\begin{pmatrix} 1+\beta^2 & -\beta \\ -\beta & 1 \end{pmatrix}$, $\mathbf{a}_0 = (\alpha_0, 0)^\top$ *e* $\mathbf{b}_0 = (\beta_0, 1)^\top$.

*ii) Para $t_{i0} = t_i(\boldsymbol{\theta}_0)$ e $s_{i0} = s_i(\boldsymbol{\theta}_0)$, temos que*

$$\begin{pmatrix} t_{i0} \\ s_{i0} \end{pmatrix} \stackrel{\text{iid}}{\sim} E\ell_2(\mathbf{0}, \mathbf{A}_0\mathbf{A}_0^\top). \tag{3.20}$$

*iii)* $(s_{i0}|t_{i0}) \sim E\ell_1\left(-\frac{\beta_0}{1+\beta_0^2}t_{i0}, \frac{1}{(1+\beta_0^2)^2}; f_{q(t_{i0})}\right)$, *onde* $q(t_{i0}) = t_{i0}^2$ *e* $f_a$ *representa a densidade condicional no caso esférico.*

**Prova:** A prova de ii) segue de i), fazendo $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ e a demonstração de iii) é conseqüência imediata das propriedades da distribuição elíptica condicional (Teorema A.4 do Apêndice A).

Para provar i), observemos que

$$\begin{pmatrix} t_i \\ s_i \end{pmatrix} = \begin{pmatrix} \mathbf{b}_2^\top \\ \mathbf{b}_1^\top \end{pmatrix}(\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_{i0}) = \mathbf{A}(\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_{i0}), \tag{3.21}$$

e

$$\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_{i0} \sim E\ell_2((\mathbf{a}_0 - \mathbf{a}) + (\mathbf{b}_0 - \mathbf{b})x_{i0}, \mathbf{I}_p), \tag{3.22}$$

então, a demonstração segue das propriedades das distribuições elípticas (veja Apêndice A).

□

De iii) do Lema 3.1 e do Teorema A.4 no Apêndice A, segue que

$$\mathrm{Var}[s_{i0}|t_{i0}] = \frac{1}{(1+\beta_0^2)^2}a(t_{i0}^2). \tag{3.23}$$

Por exemplo, se $(t_{i0}, s_{i0})^\top \sim t_2(\mathbf{0}, \mathbf{A}_0\mathbf{A}_0^\top; \nu)$, $\nu > 1$, distribuição $t$-Student, então

$$a(t_{i0}^2) = \frac{\nu + t_{i0}^2}{\nu - 1}. \tag{3.24}$$

Note também que de iii) do Lema 3.1, segue que os vetores $(t_{i0}, s_{i0})^\top$, $i = 1, \ldots, n$, são independentes e identicamente distribuídos, de modo que

$$\begin{pmatrix} t_{i0} \\ s_{i0} \end{pmatrix} \stackrel{d}{=} \begin{pmatrix} t \\ s \end{pmatrix} \sim E\ell_2(\mathbf{0}, \mathbf{A}_0\mathbf{A}_0^\top), \tag{3.25}$$

$$t_{i0} \stackrel{d}{=} t \sim E\ell_2(0,1), \quad s \sim E\ell_1\left(0, \frac{1}{||\mathbf{b}_0||^2}\right). \tag{3.26}$$

### 3.3.1 Estimação por máxima verossimilhança

Nesta seção estudaremos as propriedades assintóticas dos estimadores de máxima verossimilhança dos parâmetros estruturais do modelo funcional definido em (3.5), com $q = 1$. Para este estudo será utilizado o trabalho de Mak (1982), resumido na Seção 3.2.

Como o logaritmo da função de verossimilhança de uma amostra de tamanho $n$, definido em (3.17) depende dos parâmetros incidentais $x_1, \ldots, x_n$, primeiramente vamos estimar $x_i$, $i = 1, \ldots, n$, dado $\boldsymbol{\theta} = (\alpha, \beta)^\top$ para, posteriormente, substituí-lo em (3.17), obtendo assim uma função que depende somente da amostra e do parâmetro de interesse, como é mostrado a seguir.

**Lema 3.2** *Considere o modelo definido em (3.16) satisfazendo a condição A1. Então, o EMV condicional de $x_i$ dado $\boldsymbol{\theta}$ é dado por*

$$\tilde{x}_i = \frac{\mathbf{b}^\top}{\|\mathbf{b}\|^2}(\mathbf{Z}_i - \mathbf{a}) \tag{3.27}$$

*e*

$$h_i(\mathbf{Z}_i; \boldsymbol{\theta}) = \log f(t_i^2), \tag{3.28}$$

*onde $t_i$ é como em (3.19), $i = 1, \ldots, n$.*

**Prova:** De A1 e (3.17), segue que maximizar $\log f(\|\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_i\|^2)$ com respeito a $x_i$ dado $\boldsymbol{\theta}$ é equivalente a minimizar $\|\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_i\|^2$ com respeito a $x_i$ dado $\boldsymbol{\theta}$, de onde tem-se (3.27). Para obter (3.28), é suficiente substituir (3.27) em (3.17), obtendo-se $\|\mathbf{Z}_i - \mathbf{a} - \mathbf{b}\tilde{x}_i\|^2 = t_i^2$ e assim (3.28) é provado. □

Sejam as funções contínuas

$$W_f(u) = f'(u)/f(u) \quad \text{e} \quad \Delta_f(u) = W_{f'}(u)W_f(u) - W_f^2(u), \tag{3.29}$$

onde $f'(u)$ e $f''(u)$, $u \geq 0$, são a primeira e segunda derivadas de $f(u)$ e seja $\boldsymbol{\theta} = (\theta_1, \theta_2)^\top =$

$(\alpha, \beta)^\top \in \Theta$. Logo, das relações em (3.9), (3.10), (3.11) e (3.28), segue que

$$q_{i\theta_j}(\mathbf{Z}_i; \boldsymbol{\theta}) = W_f(t_i^2)\frac{\partial t_i^2}{\partial \theta_j}, \quad j = 1, 2, \tag{3.30}$$

$$q_{i\theta_k\theta_j}(\mathbf{Z}_i; \boldsymbol{\theta}) = \Delta_f(t_i^2)\left(\frac{\partial t_i^2}{\partial \theta_k}\right)\left(\frac{\partial t_i^2}{\partial \theta_j}\right) + W_f(t_i^2)\left(\frac{\partial^2 t_i^2}{\partial \theta_k \partial \theta_j}\right), \quad i, j = 1, 2, \tag{3.31}$$

$$q_{i\theta_k,\theta_j}(\mathbf{Z}_i; \boldsymbol{\theta}) = W_f^2(t_i^2)\left(\frac{\partial t_i^2}{\partial \theta_k}\right)\left(\frac{\partial t_i^2}{\partial \theta_j}\right), \quad i, j = 1, 2, \tag{3.32}$$

onde, depois de algumas manipulações algébricas, temos que

$$\begin{aligned}
\frac{\partial t_i^2}{\partial \alpha} &= -2(\mathbf{e}_1^\top \mathbf{b}_2)t_i, \\
\frac{\partial t_i^2}{\partial \beta} &= -2x_{i0}(\mathbf{e}_1^\top \mathbf{b}_2)t_i - \frac{2\beta}{1+\beta^2}t_i^2 - 2t_i s_i, \\
\frac{\partial^2 t_i^2}{\partial \alpha \partial \alpha} &= 2(\mathbf{e}_1^\top \mathbf{b}_2)^2, \\
\frac{\partial^2 t_i^2}{\partial \alpha \partial \beta} &= 2x_{i0}(\mathbf{e}_1^\top \mathbf{b}_2)^2 - \frac{4\beta}{1+\beta^2}(\mathbf{e}_1^\top \mathbf{b}_2)t_i + 2(\mathbf{e}_1^\top \mathbf{b}_2)s_i \\
\frac{\partial^2 t_i^2}{\partial \beta \partial \beta} &= 2x_{i0}^2(\mathbf{e}_1^\top \mathbf{b}_2)^2 + 4x_{i0}(e_1^\top b_2)\left(\frac{2\beta}{1+\beta^2}t_i + s_i\right) \\
&\quad + 2\frac{3\beta^2 - 1}{(1+\beta^2)^2}t_i^2 + \frac{8\beta}{2+\beta^2}t_i s_i + 2s_i^2.
\end{aligned}$$

**Lema 3.3** *Considere o modelo definido em (3.16) satisfazendo as condições A1 e A2. Então as matrizes $E_0[A_n(\boldsymbol{\theta}_0)]$ e $V_n(\boldsymbol{\theta}_0)$, definidas em (3.9) e (3.13), respectivamente, são dadas por*

$$E_0[A_n(\boldsymbol{\theta}_0)] = \frac{2}{1+\beta_0^2}C_f \begin{pmatrix} 1 & \overline{x}_0 \\ \overline{x}_0 & S_x^0 + \frac{1}{1+\beta_0^2}B_f/C_f \end{pmatrix}$$

*e*

$$V_n(\boldsymbol{\theta}_0) = \frac{4}{1+\beta_0^2}E_0[W_f^2(t^2)t^2] \begin{pmatrix} 1 & \overline{x}_0 \\ \overline{x}_0 & S_x^0 + \frac{1}{1+\beta_0^2}A_f \end{pmatrix},$$

*onde*

$$\begin{aligned}
\overline{x}_0 &= \frac{1}{n}\sum_{i=1}^n x_{i0}, \quad S_x^0 = \frac{1}{n}\sum_{i=1}^n x_{i0}^2, \\
A_f &= \frac{E_0[W_f(t^2)a(t^2)t^2]}{E_0[W_f^2(t^2)t^2]},
\end{aligned}$$

$$B_f = E_0[W_f(t^2)(a(t^2) - t^2)] + 2E_0[\Delta_f(t^2)a(t^2)t^2]$$

$$\text{e} \qquad C_f = E_0[W_f(t^2)] + 2E_0[\Delta_f(t^2)t^2],$$

*com $a(t^2)$ e $(t,s)^\top$ sendo como em (3.23) e (3.25). respectivamente.*

**Prova:** De (3.31) e do fato que $t \stackrel{\mathrm{d}}{=} t_{i0}$, temos que

$$E_0[q_{i\alpha\alpha}(\mathbf{Z}_i;\boldsymbol{\theta}_0)] = \frac{2}{1+\beta_0^2}\left\{E_0[W_f(t^2)] + 2E_0[\Delta_f(t^2)t^2]\right\} = \frac{2}{1+\beta_0^2}C_f.$$

Agora, do fato que (veja Lema 3.1)

$$E_0[t_{i0}] = 0 \quad \text{e} \quad E_0[s_{i0}|t_{i0}] = -\frac{\beta_0}{1+\beta_0^2}t_{i0}, \tag{3.33}$$

segue que

$$E_0[q_{i\alpha\beta}(\mathbf{Z}_i;\boldsymbol{\theta}_0)] = x_{i0}E_0[q_{i\alpha\alpha}(\mathbf{Z}_i;\boldsymbol{\theta}_0)].$$

Similarmente, de (3.31), (3.23) e (3.33), e após algumas manipulações algébricas, temos que

$$\begin{aligned} E_0[q_{i\beta\beta}(\mathbf{Z}_i;\boldsymbol{\theta}_0)] &= x_{i0}^2E_0[q_{i\alpha\alpha}(\mathbf{Z}_i;\boldsymbol{\theta}_0)] \\ &\quad + \frac{2}{(1+\beta_0^2)}\left\{E_0[W_f(t^2)(a(t^2)-t^2)] + E_0[\Delta_f(t^2)a(t^2)t^2]\right\} \\ &= x_{i0}^2E_0[q_{i\alpha\alpha}(\mathbf{Z}_i;\boldsymbol{\theta}_0)] + \frac{2}{(1+\beta_0^2)^2}B_f. \end{aligned}$$

Assim, somando as expressões acima em $i = 1,\ldots,n$ e dividindo por $n$, temos de (3.12) a matriz $E_0[A_n(\boldsymbol{\theta}_0)]$.

Por outro lado, considerando (3.33) e a simetria da distribuição de $(t_{i0}, s_{i0})^\top$, segue de (3.30) que

$$E_0[q_{i\alpha}(\mathbf{Z}_i;\boldsymbol{\theta}_0)] = E_0[q_{i\beta}(\mathbf{Z}_i;\boldsymbol{\theta}_0)] = 0. \tag{3.34}$$

Assim, para determinar a matriz $V_n(\boldsymbol{\theta}_0)$ definida em (3.13) é suficiente calcular os termos $E_0[q_{i\theta_k,\theta_j}(\mathbf{Z}_i;\boldsymbol{\theta}_0)]$, $j,k = 1,2$, que são obtidos de forma similar às entradas da matriz

$E_0[A_i(\mathbf{Z}_i;\boldsymbol{\theta}_0)]$. Desta forma, temos que

$$E_0[q_{i\alpha,\alpha}(\mathbf{Z}_i;\boldsymbol{\theta}_0)] = \frac{4}{1+\beta_0^2}E_0[W_f^2(t^2)t^2],$$

$$E_0[q_{i\alpha,\beta}(\mathbf{Z}_i;\boldsymbol{\theta}_0)] = x_{i0}E_0[q_{i\alpha,\alpha}(\mathbf{Z}_i;\boldsymbol{\theta}_0)]$$

e $$E_0[q_{i\beta,\beta}(\mathbf{Z}_i;\boldsymbol{\theta}_0)] = x_{i0}^2E_0[q_{i\alpha,\alpha}(\mathbf{Z}_i;\boldsymbol{\theta}_0)] + \frac{4}{(1+\beta_0^2)^2}E_0[W_f^2(t^2)a(t^2)t^2].$$

Somando agora as expressões acima em $i$, $i = 1, 2, \ldots, n$, e dividindo por $n$, temos a matriz $V_n(\boldsymbol{\theta}_0)$. □

Para provar os principais resultados desta seção, as seguintes condições de regularidade são requeridas:

A3. Existem constantes positivas $a_f$ e $b_f$ tais que

$$|W_f(u)| \leq a_f \quad \text{e} \quad |W_{f'}(u)| \leq b_f, \quad u \geq 0.$$

A4. Para cada $\overline{\boldsymbol{\theta}}$ no interior de $\Theta$, existe $\delta > 0$ e funções $d_i(\mathbf{Z}_i)$ e $d_{ijk}(\mathbf{Z}_i)$ tais que

$$|h_i(\mathbf{Z}_i;\boldsymbol{\theta})| \leq d_i(\mathbf{Z}_i) \quad \text{e} \quad |q_{i\theta_k\theta_j}(\mathbf{Z}_i;\boldsymbol{\theta})| \leq d_{ijk}(\mathbf{Z}_i),$$

para todo $\boldsymbol{\theta} \in B(\overline{\boldsymbol{\theta}},\delta) = \{\boldsymbol{\theta} : \|\overline{\boldsymbol{\theta}} - \boldsymbol{\theta}\| < \delta\} \subset \Theta^0$ e

$$\limsup n^{-1}\sum_{i=1}^{n} E_0[d_i^2(\mathbf{Z}_i)] < \infty, \qquad \limsup n^{-1}\sum_{i=1}^{n} E_0[d_{ijk}^2(\mathbf{Z}_i;\theta)] < \infty.$$

A5. Existe $\gamma > 0$ tal que

$$E_0[|W_f(t^2)t^2|^{2+\gamma}] < \infty \quad \text{e} \quad E_0[|W_f(t^2)ts|^{2+\gamma}] < \infty.$$

A6. A seqüência $x_1, x_2, \ldots$ é limitada, isto é, existe $M > 0$ tal que $\sup_{i\geq 1}|x_i| \leq M$.

A7. Dado $\varepsilon > 0$, existe $\delta > 0$ tal que

$$\limsup \frac{1}{n}\left|\sum_{i=1}^{n} E_0\left[\sup\left\{q_{i\theta_k\theta_j}(\mathbf{Z}_i;\boldsymbol{\theta}) : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| < \delta\right\} - q_{i\theta_k\theta_j}(\mathbf{Z}_i;\boldsymbol{\theta}_0)\right]\right| < \varepsilon,$$

$k, j = 1, 2$. O mesmo é verdadeiro quando o supremo é substituído pelo ínfimo.

Observações:

1. Em certas situações, a condição A6 pode ser substituída por uma condição mais fraca, a saber

   (a) $0 < \liminf S_{xx} \leq \limsup S_{xx} < \infty$, onde

   $$S_{xx} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 \quad \text{e} \quad \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

   (b) Para $\gamma$ satisfazendo a condição A5, temos que

   $$\lim_n \frac{1}{n^{1+\gamma/2}} \sum_{i=1}^{n} |x_i|^{2+\gamma} = 0.$$

   Tal fato ocorre, por exemplo, no modelo normal (veja Mak, 1982).

2. Se existe a terceira derivada de $f(u)$, $u \geq 0$, então a condição A7 pode ser substituída pela seguinte condição:

   A7'. Existe uma função positiva $d_{iklm}(\mathbf{Z}_i)$ tal que

   $$\left|\frac{\partial}{\partial \theta_m} q_{i\theta_k\theta_l}(\mathbf{Z}_i; \boldsymbol{\theta})\right| \leq d_{iklm}(\mathbf{Z}_i),$$

   para $\boldsymbol{\theta}$ numa vizinhança de $\boldsymbol{\theta}_0$ e

   $$\limsup \frac{1}{n}\sum_{i=1}^{n} E_0[d_{iklm}(\mathbf{Z}_i)] < \infty, \quad k, l, m = 1, 2.$$

Com a finalidade de provar que $\boldsymbol{\theta}_0$ é um ponto de máximo local da função $\overline{\psi}(\boldsymbol{\theta})$, definida em (3.14), precisamos de um resultado sobre pontos críticos de funções definida em $\mathbb{R}^n$, que apresentamos a seguir.

Seja $\varphi$ de classe $C^{(2)}$ (ou $\varphi \in C^{(2)}$) sobre um conjunto aberto $D \subseteq \mathbb{R}^m$ e consideremos a função $Q_\varphi : \mathbf{D} \times \mathbb{R}^m \to \mathbb{R}$ definida por

$$Q_\varphi(\mathrm{x}, \mathrm{h}) = \sum_{jk} \varphi_{jk}(\mathrm{x}) h_j h_k = \mathrm{h}^\top \boldsymbol{\Psi} \mathrm{h},$$

onde $\Psi = (\varphi_{jk})$ e os $\varphi_{jk}$ são as segundas derivadas parciais de $\varphi$ e $\mathbf{h} \in \mathbb{R}^m$. Assim, para cada $\mathbf{x}_0 \in \mathbf{D}$, a função $Q_\varphi : \mathbb{R}^m \to \mathbb{R}$ definida por $Q_\varphi^{x_0}(\mathbf{h}) = Q_\varphi(\mathbf{x}_0, \mathbf{h})$ é a forma quadrática correspondente à matriz $(\varphi_{jk}(\mathbf{x}_0))$ das segundas derivadas. Diremos que a matriz $(\varphi_{jk}(\mathbf{x}_0))$ é negativa definida se $Q_\varphi(\mathbf{x}_0, \mathbf{h}) < 0$, para todo $\mathbf{h} \in \mathbb{R}^m$, $\mathbf{h} \neq 0$, que denotamos por $Q_\varphi(\mathbf{x}_0, \cdot) < 0$. Se $Q_\varphi(\mathbf{x}_0, \mathbf{h}) > 0$, para todo $\mathbf{h} \in \mathbb{R}^m$, $\mathbf{h} \neq 0$, então diremos que a matriz $(\varphi_{jk}(\mathbf{x}_0))$ é positiva definida.

Sejam $\lambda_1(\mathbf{x}_0), \ldots, \lambda_m(\mathbf{x}_0)$ os autovalores da matriz $(\varphi_{jk}(\mathbf{x}_0))$. Então $Q_\varphi(\mathbf{x}_0, \cdot) < 0$ $(> 0)$ se e somente se $\lambda_i(\mathbf{x}_0) < 0$ $(> 0)$, $i = 1, \ldots, m$. O resultado anterior e o lema seguinte podem ser encontrados, por exemplo, em Fleming (1966).

**Lema 3.4** *Seja $\varphi \in C^{(2)}$ sobre um conjunto aberto convexo $\mathbf{D}$, e seja $\mathbf{x}_0 \in \mathbf{D}$ um ponto crítico de $\varphi$. Se $Q_\varphi(\mathbf{x}_0, \cdot) < 0$, então $\mathbf{x}_0$ é um ponto de máximo local da função $\varphi$.*

O primeiro resultado importante que daremos a seguir mostra que o ponto $\boldsymbol{\theta}_1$ (definido na Seção 3.2) coincide com o verdadeiro valor $\boldsymbol{\theta}_0$, sob condições adicionais. Para provar este resultado, o Lema 3.4 será de grande utilidade. Este resultado requer condições sobre os coeficientes $B_f$ e $C_f$ definidos no Lema 3.3. Sob normalidade $B_f = 0$ e $C_f < 0$ (veja Mak, 1982). Tais propriedades ainda valem no modelo $t$-Student como será mostrado na Seção 3.4. Conjecturamos que na família das distribuições elípticas também valem estes resultados, pois da condição A1 a função de densidade $f(\mathbf{x}^\top \mathbf{x})$, $\mathbf{x} \in \mathbb{R}^p$, têm um comportamento simétrico similar à densidade do modelo normal (simetria com relação ao parâmetro de locação).

**Teorema 3.1** *Suponhamos que $B_f = 0$, $C_f < 0$ e que as condições A1 - A3 estejam satisfeitas. Então, $\boldsymbol{\theta}_0$ é um ponto de máximo local da função $\overline{\psi}(\boldsymbol{\theta}) = n^{-1} \sum_{i=1}^n E_0[h_i(\mathbf{Z}_i; \boldsymbol{\theta})]$, onde $h_i(\mathbf{Z}_i; \boldsymbol{\theta}) = \log f(t_i^2)$, $i = 1, \ldots, n$.*

**Prova:** Para cada $i$ e $\boldsymbol{\theta} \in \Theta$, segue de (3.30) que

$$q_{i\alpha}(\mathbf{Z}_i; \boldsymbol{\theta}) = -2(\mathbf{e}_1^\top \mathbf{b}_2) W_f(t_i^2) t_i$$

e

$$q_{i\beta}(\mathbf{Z}_i;\boldsymbol{\theta}) = x_{i0}q_{i\alpha}(\mathbf{Z}_i;\boldsymbol{\theta}) - \frac{2}{\|\mathrm{b}\|}W_f(t_i^2)\left(\frac{\beta}{1-\beta^2}t_i^2 + t_is_i\right).$$

Agora, para $\boldsymbol{\theta} \in B(\boldsymbol{\theta}_0, \delta)$, $\delta > 0$, temos que

$$\|\mathbf{Z}_i - \mathrm{a} - \mathrm{b}x_{i0}\| \leq \|\mathbf{Z}_i - \mathrm{a}_0 - \mathrm{b}_0x_{i0}\| + \delta(1 + |x_{i0}|) = d_0(\mathbf{Z}_i).$$

Assim, pela desigualdade de Cauchy-Schwarz, segue que

$$|t_i| \leq d_0(\mathbf{Z}_i) \quad \text{e} \quad |s_i| \leq d_0(\mathbf{Z}_i)$$

e, conseqüentemente, da condição A3, obtemos

$$|q_{i\alpha}(\mathbf{Z}_i;\boldsymbol{\theta})| \leq 2a_fd_0(\mathbf{Z}_i)$$

e

$$|q_{i\beta}(\mathbf{Z}_i;\boldsymbol{\theta})| \leq a_f(2d_0(\mathbf{Z}_i)|x_{i0}| + 3d_0^2(\mathbf{Z}_i)).$$

Definindo

$$\psi_i(\boldsymbol{\theta}) = E_0[h_i(\mathbf{Z}_i;\boldsymbol{\theta})], \quad i = 1, \ldots, n,$$

segue das desigualdades acima, da condição A2 e do Teorema da Convergência Dominada, que

$$\frac{\partial\psi_i(\boldsymbol{\theta})}{\partial\theta_j} = E_0[q_{i\theta_j}(\mathbf{Z}_i;\boldsymbol{\theta})], \quad j = 1, 2.$$

Assim, de (3.34), temos que $\boldsymbol{\theta}_0$ é um ponto crítico da função $\psi_i(\boldsymbol{\theta})$, $i = 1, \ldots, n$, e conseqüentemente um ponto crítico de $\overline{\psi}(\boldsymbol{\theta})$. Por outro lado, da desigualdade de Cauchy-Schwarz segue que, para $\boldsymbol{\theta} \in B(\boldsymbol{\theta}_0, \delta)$,

$$|q_{i\alpha\alpha}(\mathbf{Z}_i;\boldsymbol{\theta})| \leq 4c_fd_0^2(\mathbf{Z}_i) + 2a_f\ ,$$

$$|q_{i\alpha\beta}(\mathbf{Z}_i;\boldsymbol{\theta})| \leq 2c_f[2d_0^2(\mathbf{Z}_i)|x_{i0}| + 3d_0^3(\mathbf{Z}_i)] + 2a_f[|x_{i0}| + 2d_0(\mathbf{Z}_i)]$$

e $$|q_{i\beta\beta}(\mathbf{Z}_i;\boldsymbol{\theta})| \leq 2c_f[2d_0(\mathbf{Z}_i)|x_{i0}| + 3d_0^2(\mathbf{Z}_i)]^2 + 2a_f[|x_{i0}^2| + 4|x_{i0}|d_0(\mathbf{Z}_i) + 6d_0^2(\mathbf{Z}_i)]\ ,$$

onde $c_f = a_f(a_f + b_f)$. Assim, da condição A2 e do Teorema da Convergência Dominada, temos que

$$\frac{\partial^2 \psi_i(\boldsymbol{\theta})}{\partial \theta_k \partial \theta_j} = E_0[q_{i\theta_k\theta_j}(\mathbf{Z}_i; \boldsymbol{\theta})], \quad k, j = 1, 2,$$

e $\psi_i \in C^{(2)}$, $i = 1, \ldots, n$, pois $f \in C^2$. Além disso, a matriz das segundas derivadas de $\psi_i(\boldsymbol{\theta})$ avaliada em $\boldsymbol{\theta}_0$ é dada pela matriz $E_0[A_i(\mathbf{Z}_i; \boldsymbol{\theta}_0)]$, definida por

$$E_0[A_i(\mathbf{Z}_i; \boldsymbol{\theta}_0)] = \frac{2C_f}{1+\beta_0^2} \begin{pmatrix} 1 & x_{i0} \\ x_{i0} & x_{i0}^2 + \frac{1}{1+\beta_0^2} B_f / C_f \end{pmatrix}, \quad i = 1, \ldots, n.$$

Daí, segue que a matriz das segundas derivadas de $\overline{\psi}(\boldsymbol{\theta}) = n^{-1} \sum_{i=1}^{n} \psi_i(\boldsymbol{\theta})$ é dada pela matriz $E_0[A_n(\boldsymbol{\theta}_0)]$ definida no Lema 3.3 com $B_f = 0$, isto é,

$$E_0[A_n(\boldsymbol{\theta}_0)] = \frac{2C_f}{1+\beta_0^2} \begin{pmatrix} 1 & \overline{x}_0 \\ \overline{x}_0 & S_x^0 \end{pmatrix},$$

cujos autovalores são dados por

$$\lambda_1 = \frac{C_f}{1+\beta_0^2} \left[ (1 + S_{x_0}) + \left\{ (1 + S_{x_0})^2 - 4(S_{x_0} - \overline{x}_0^2) \right\}^{1/2} \right]$$

e

$$\lambda_2 = \frac{C_f}{1+\beta_0^2} \left[ (1 + S_{x_0}) - \left\{ (1 + S_{x_0})^2 - 4(S_{x_0} - \overline{x}_0^2) \right\}^{1/2} \right].$$

Notemos que

$$(1 + S_{x_0})^2 - 4(S_{x_0} - \overline{x}_0^2) = (1 - S_{x_0})^2 + 4x_0^2 > 0$$

e

$$1 + S_{x_0} > [(1 + S_{x_0})^2 - 4(S_{x_0} - \overline{x}_0^2)]^{1/2},$$

o que implica que $\lambda_1 < 0$ e $\lambda_2 < 0$, pois $C_f < 0$. Assim, do Lema 3.4, segue que a matriz $E_0[A_n(\boldsymbol{\theta}_0)]$ é negativa definida e, portanto, $\boldsymbol{\theta}_0$ é um máximo local de $\overline{\psi}(\boldsymbol{\theta})$. □

O resultado que apresentamos a seguir mostra a existência de um estimador consistente e assintoticamente normal que maximiza a função de verossimilhança em (3.7). Este estimador corresponde ao EMV $\widetilde{\boldsymbol{\theta}}_n$ que é obtido resolvendo-se simultaneamente as equações (veja Mak, 1982)

$$\sum_{i=1}^{n} q_{i\alpha}(\mathbf{Z}_i; \boldsymbol{\theta}) = 0 \quad \text{e} \quad \sum_{i=1}^{n} q_{i\beta}(\mathbf{Z}_i; \boldsymbol{\theta}) = 0 \ . \tag{3.35}$$

**Teorema 3.2** *Considere o modelo definido em (3.16) satisfazendo as condições A1 - A7, com* $B_f = 0$ *e* $C_f < 0$. *Então, existe um estimador* $\widetilde{\boldsymbol{\theta}}_n = (\widetilde{\alpha}_n, \widetilde{\beta}_n)^\top$ *de* $\boldsymbol{\theta}_0$ *que maximiza (3.7), o qual é obtido resolvendo-se a equação (3.35) tal que* $\widetilde{\boldsymbol{\theta}}_n \xrightarrow{\mathrm{P}} \boldsymbol{\theta}_0$. *Além disso,* $\widetilde{\boldsymbol{\theta}}_n$ *é assintoticamente normal com média* $\boldsymbol{\theta}_0$ *e matriz de covariâncias dada por*

$$\boldsymbol{\Sigma}_n = \frac{E_0[W_f^2(t^2)t^2]}{C_f^2(S_{xx}^0)^2} \begin{pmatrix} \Delta(S_{xx}^0)^2 + K\overline{x}_0^2 & -K\overline{x}_0 \\ -K\overline{x}_0 & K \end{pmatrix}, \tag{3.36}$$

*onde* $\Delta = 1 + \beta_0^2$ *e* $K = \Delta S_{xx}^0 + A_f$, *com* $A_f$ *e* $C_f$ *sendo como no Lema 3.3.*

**Prova:** Utilizando a desigualdade de Cauchy-Schwarz, temos sob as condições A1 – A7 que o modelo funcional elíptico satisfaz as condições de regularidade definidas em Mak (1982) (veja Seção 2). Assim, segue dos resultados da Seção 3.2 (veja Mak, 1982) que existe um estimador $\widetilde{\boldsymbol{\theta}}_n = (\widetilde{\alpha}_n, \widetilde{\beta}_n)^\top$ consistente para $\boldsymbol{\theta}_0$ que maximiza (3.7) e que é assintoticamente normal com média $\boldsymbol{\theta}_0$ e matriz de covariância assintótica dada por

$$\boldsymbol{\Sigma}_n = (E_0[A_n(\boldsymbol{\theta}_0)])^{-1} V_n(\boldsymbol{\theta}_0)(E_0[A_n(\boldsymbol{\theta}_0)])^{-1},$$

onde

$$V_n(\boldsymbol{\theta}_0) = \frac{4}{1+\beta_0^2} E_0[W_f^2(t^2)t^2] \begin{pmatrix} 1 & \overline{x}_0 \\ \overline{x}_0 & S_x^0 + A_f \end{pmatrix}$$

e

$$(E_0[A_n(\boldsymbol{\theta}_0)])^{-1} = \frac{1+\beta_0^2}{2C_f S_{xx}^0} \begin{pmatrix} S_x^0 & -\overline{x}_0 \\ -\overline{x}_0 & 1 \end{pmatrix}.$$

Assim, após algumas manipulações algébricas, temos (3.36). □

### 3.3.2 Eficiência relativa assintótica

Para efeito de estudar a eficiência relativa assintótica do EMV de $\beta_0$, com respeito ao estimador de mínimos quadrados generalizados (Sprent, 1966, Gleser, 1985), que resulta em minimizar a função

$$Q_G(\alpha_0, \beta_0, \mathrm{x}_0) = \sum_{i=1}^{n} (\mathbf{Z}_i - \mathrm{a}_0 - \mathrm{b}_0 x_{i0})^\top (\mathbf{Z}_i - \mathrm{a}_0 - \mathrm{b}_0 x_{i0}), \tag{3.37}$$

onde $\mathrm{x}_0 = (x_{i0}, \ldots, x_{n0})^\top$ e que é dado por

$$\widehat{\beta}_{GLS} = \frac{S_{YY} - S_{XX} + \sqrt{(S_{YY} - S_{XX})^2 + 4S_{XY}^2}}{2S_{XY}}, \tag{3.38}$$

onde $S_{YY}$, $S_{XX}$ e $S_{XY}$ são as entradas da matriz de covariâncias amostrais $\mathbf{S}_n$, é necessário supor condições adicionais sobre os parâmetros incidentais $x_{10}, \ldots, x_{n0}$, a saber:

A8. existem constantes $\mu_x$ e $\sigma^*_{xx} > 0$ tais que $\lim_n \overline{x}_0 = \mu_x$ e $\lim S^0_{xx} = \sigma^*_{xx}$.

O resultado que segue será de grande utilidade para o estudo desta eficiência assintótica relativa dos estimadores $\widehat{\beta}_n$ e $\widehat{\beta}_{GLS}$ que denotamos por $e_{\widehat{\beta}_{GLS}, \widetilde{\beta}_n}$

**Teorema 3.3** *Sob as condições do Teorema 3.2 e sendo a condição A8 satisfeita, tem-se que o EMV* $\widetilde{\boldsymbol{\theta}}_n = (\widetilde{\alpha}_n, \widetilde{\beta}_n)^\top$ *de* $\boldsymbol{\theta}_0 = (\alpha_0, \beta_0)^\top$ *é tal que* $\widetilde{\boldsymbol{\theta}}_n \xrightarrow{\mathrm{P}} \boldsymbol{\theta}_0$. *Além disso,* $\widetilde{\boldsymbol{\theta}}_n$ *é assintoticamente normal com média* $\boldsymbol{\theta}_0$ *e matriz de covariâncias dada por*

$$\mathbf{\Sigma}_* = \frac{E[W_f^2(t^2)t^2]}{C_f^2(\sigma^*_{xx})^2} \begin{pmatrix} \Delta(\sigma^*_{xx})^2 + K\mu_x^2 & -K\mu_x \\ -K\mu_x & K \end{pmatrix}, \tag{3.39}$$

*onde* $K$ *e* $\Delta$ *são como no Teorema 3.2.*

**Prova:** Do Teorema 3.2, temos que existe um estimador $\widetilde{\boldsymbol{\theta}}_n$ de $\boldsymbol{\theta}_0$ tal que $\widetilde{\boldsymbol{\theta}}_n \xrightarrow{\mathrm{P}} \boldsymbol{\theta}_0$ e

$$(V_n(\boldsymbol{\theta}_0))^{-1/2}(E_0[A_n(\boldsymbol{\theta}_0)])\sqrt{n}(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{\mathrm{d}} N_2(0, I_2), \tag{3.40}$$

onde $V_n(\boldsymbol{\theta}_0)$ e $E_0[A_n(\boldsymbol{\theta}_0)]$ são como no Lema 3.3, com $B_f = 0$. Agora, pela condição A8, temos que

$$V_n(\boldsymbol{\theta}_0) \xrightarrow{\mathrm{P}} V(\boldsymbol{\theta}_0) \quad \text{e} \quad E_0[A_n(\boldsymbol{\theta}_0)] \xrightarrow{\mathrm{P}} E_0[A(\boldsymbol{\theta}_0)], \tag{3.41}$$

onde

$$V(\boldsymbol{\theta}_0) = \frac{4}{1+\beta_0^2} E[W_f^2(t^2)t^2] \begin{pmatrix} 1 & \mu_x \\ \mu_x & \sigma_{xx}^* + \mu_x + \frac{1}{1+\beta_0^2} A_f \end{pmatrix}$$

e

$$E[A(\boldsymbol{\theta}_0)] = \frac{2C_f}{1+\beta_0^2} \begin{pmatrix} 1 & \mu_x \\ \mu_x & \sigma_{xx}^* + \mu_x \end{pmatrix}.$$

Portanto, de (3.40), (3.41) e do Teorema de Slutsky, tem-se que

$$(V(\boldsymbol{\theta}_0)^{-1/2}(E_0[A(\boldsymbol{\theta}_0)])\sqrt{n}(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{\mathrm{d}} N_2(\mathbf{0}, \mathbf{I}_2), \tag{3.42}$$

isto é, $\widetilde{\boldsymbol{\theta}}_n$ é assintoticamente normal com média $\boldsymbol{\theta}_0$ e matriz de covariâncias dada por

$$\boldsymbol{\Sigma}_* = (E_0[A(\boldsymbol{\theta}_0)])^{-1} V(\boldsymbol{\theta}_0)(E_0[A(\boldsymbol{\theta}_0)])^{-1}.$$

Após algumas manipulações algébricas, tem-se que $\boldsymbol{\Sigma}_*$ é como em (3.39). □

Note que a matriz $\boldsymbol{\Sigma}_*$ depende dos parâmetros $\mu_x$, $\sigma_{xx}^*$ e $\beta_0$, isto é, $\boldsymbol{\Sigma}_* = \boldsymbol{\Sigma}_*(\mu_x, \sigma_{xx}^*, \beta_0)$. Logo, para estimar $\boldsymbol{\Sigma}_*$ consistentemente é suficiente estimar consistentemente $\mu_x$, $\sigma_{xx}^*$ e $\beta_0$. Seja $\mathbf{S}_n$ a matriz de covariâncias amostrais e sejam as variáveis aleatórias

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{e} \quad \widehat{S}_0 = \frac{1}{\|\mathbf{b}_0\|^2}\left(\frac{\mathbf{b}_0^\top \mathbf{S}_n \mathbf{b}_0}{\|\mathbf{b}_0\|^2} - 1\right). \tag{3.43}$$

Então, de i) do Teorema 2.2, temos que

$$\lim_n \overline{X} = \mu_x \quad \text{e} \quad \lim_n \widehat{S}_0 = \sigma_{xx}^*, \tag{3.44}$$

pois do Teorema 2.2, $\lim \mathbf{S}_n = \mathbf{I}_2 + \sigma_{xx}^* \mathbf{b}_0 \mathbf{b}_0^\top$. Assim, temos que $\overline{X}$ e $\widehat{S}_0$ são estimadores consistentes de $\mu_x$ e $\sigma_{xx}^*$, respectivamente.

**Corolário 3.1** *Sob as condições do Teorema 3.3, tem-se que*

$$e_{\widehat{\beta}_{GLS}, \widetilde{\beta}_n} = \frac{E_0[W_f^2(t^2)t^2]}{C_f^2 \delta}\left(\frac{\Delta\sigma_{xx}^* + A_f}{\Delta\sigma_{xx}^* + \delta(\kappa+1)}\right), \tag{3.45}$$

*onde*

$$\kappa = \frac{\phi''(0)}{(\phi'(0))^2} - 1$$

*e* $\delta = -2\phi'(0)$, *com* $\phi(u)$, $u \geq 0$, *sendo a função característica de* $t \sim E\ell_1(0, 1; \phi)$.

**Prova:** Do Teorema 2.3, segue que

$$\sqrt{n}(\widehat{\beta}_{GLS} - \beta_0) \xrightarrow{\mathrm{d}} N(0, \Delta_{GLS}),$$

onde

$$\Delta_{GLS} = \frac{\delta}{(\sigma_{xx}^*)^2}(\Delta\sigma_{xx}^* + \delta(\kappa + 1)). \tag{3.46}$$

Agora, do Teorema 3.3, segue que

$$\sqrt{n}(\widetilde{\beta}_n - \beta_0) \xrightarrow{\mathrm{d}} N(0, \sigma_M^2),$$

onde

$$\sigma_M^2 = \frac{E_0[W_f^2(t^2)t^2]}{C_f^2(\sigma_{xx}^*)^2}(\Delta\sigma_{xx}^* + A_f). \tag{3.47}$$

Assim, (3.45) segue de (3.46), (3.47) e do fato que $e_{\widehat{\beta}_{GLS},\widetilde{\beta}_n} = \sigma_M^2/\Delta_{GLS}$. □

Sob normalidade, Gleser (1981) prova que o EMV coincide com o estimador de Mínimos Quadrados Generalizados. Neste caso, $e_{\widehat{\beta}_{GLS},\widetilde{\beta}_n} = 1$.

### 3.3.3 O modelo funcional $t$-Student univariado

Nesta seção estudaremos o modelo funcional elíptico que resulta de considerar em (3.16), $q = 1$, e a função $f$, definida por

$$f(u) = k(2,\nu)\nu^{\nu/2}(\nu + u)^{-(\nu+2)/2}, \; u \geq 0, \tag{3.48}$$

com

$$k(2,\nu) = \frac{\Gamma(\frac{1}{2}(\nu + 2))}{\pi\Gamma(\frac{1}{2}\nu)}.$$

Em tal caso, temos o modelo funcional $t$-Student, desde que em (3.4) tem-se que $\boldsymbol{\epsilon}_1 \sim t_2(\mathbf{0}, \mathbf{I}_2; \nu)$, a distribuição $t$-Student padrão, com $\nu$ graus de liberdade.

Neste modelo, é fácil ver que

$$W_f(u) = -\frac{\nu + 2}{2}(\nu + u)^{-1}, \quad \Delta_f(u) = \frac{\nu + 2}{2}(\nu + u)^{-2} \tag{3.49}$$

e a função $h_i(\mathbf{Z}_i;\theta)$ definida em (3.28) é dada por

$$h_i(\mathbf{Z}_i;\theta) = \log(k(\nu,2)\nu^{\nu/2}(\nu+t_i^2)^{-(\nu+2)/2} = -\log 2\pi - \frac{\nu+2}{2}\log\left(1+\frac{t_i^2}{\nu}\right), \tag{3.50}$$

o que implica que

$$|h_i(\mathbf{Z}_i;\theta)| = \log 2\pi + \frac{\nu+2}{2}\log\left(1+\frac{t_i^2}{\nu}\right). \tag{3.51}$$

Aqui, temos que a seqüência $(t_{10},s_{10})^\top,\ldots(t_{n0},s_{n0})^\top$ é iid com distribuição comum

$$\begin{pmatrix} t \\ s \end{pmatrix} \sim t_2(\mathbf{0}.\,\mathbf{A}_0\mathbf{A}_0^\top;\nu), \tag{3.52}$$

sendo $\mathbf{A}_0\mathbf{A}_0^\top$ é como no Lema 3.1, de onde segue que

$$(s|t) \sim t_1\left(-\frac{\beta}{1+\beta^2}t, \frac{1}{(1+\beta^2)^2};\nu+1\right), \tag{3.53}$$

de modo que

$$\mathrm{Cov}_0[s|t] = \frac{1}{1+\beta_0^2}a(t^2), \tag{3.54}$$

onde

$$a(t^2) = \frac{\nu+t^2}{\nu-1} = -\frac{1}{2}\frac{\nu+2}{\nu-1}W_f^{-1}(t^2).$$

Para este modelo funcional $t$-Student, as matrizes $V_n(\boldsymbol{\theta}_0)$ e $E_0[A_n(\boldsymbol{\theta}_0)]$ definidas no Lema 3.3 dependem das seguintes esperanças, que foram calculadas usando o Lema A.1 do Apêndice A:

$$E_0[W_f^2(t^2)a(t^2)t^2] = \frac{1}{4}\frac{(\nu+2)^2}{(\nu-1)(\nu+1)};\quad E_0[\Delta_f(t^2)a(t^2)t^2] = \frac{1}{2}\frac{\nu+2}{(\nu-1)(\nu+1)};$$
$$E_0[W_f^2(t^2)a(t^2)] = -\frac{1}{2}\frac{\nu+2}{\nu-1};\quad E_0[W_f(t^2)t^2] = -\frac{1}{2}\frac{\nu+2}{\nu+1};$$
$$E_0[W_f^2(t^2)t^2] = \frac{1}{4}\frac{(\nu+2)^2}{(\nu+1)(\nu+3)};\quad E_0[W_f(t^2)] = -\frac{1}{2}\frac{\nu+2}{\nu+1};$$
$$\text{e}\qquad E_0[\Delta_f(t^2)t^2] = \frac{1}{2}\frac{\nu+2}{(\nu+1)(\nu+3)}.$$

Para a obtenção das esperanças acima, veja também o Lema 3.10, na Seção 3.4.2. Uma demonstração mais detalhada no caso multivariado será dada na Seção 3.4. Assim, usando as expressões dadas acima temos, após alguma manipulação algébrica, que as constantes $A_f$, $B_f$ e $C_f$, definidas no Lema 3.3, são dadas por

$$A_f = \frac{\nu+3}{\nu-1}, \quad B_f = 0 \quad \text{e} \quad C_f = -\frac{1}{2}\frac{\nu+2}{\nu+3}. \tag{3.55}$$

Conseqüentemente, as matrizes $V_n(\boldsymbol{\theta}_0)$ e $E_0[A_n(\boldsymbol{\theta}_0)]$ são dadas por

$$V_n(\boldsymbol{\theta}_0) = \frac{1}{1+\beta_0^2}\frac{(\nu+2)^2}{(\nu+1)(\nu+3)}\begin{pmatrix} 1 & \overline{x}_0 \\ \overline{x}_0 & S_x^0 + \frac{1}{1+\beta_0^2}A_f \end{pmatrix} \tag{3.56}$$

e

$$E_0[A_n(\boldsymbol{\theta}_0)] = -\frac{1}{1+\beta_0^2}\frac{\nu+2}{\nu+3}\begin{pmatrix} 1 & \overline{x}_0 \\ \overline{x}_0 & S_x^0 \end{pmatrix}. \tag{3.57}$$

**Teorema 3.4** *Suponha que o modelo funcional definido em (3.16) com $f$ dado em (3.48) satisfaz a condição A6. Então, para $\nu > 4$, o estimador de máxima verossimilhança $\widetilde{\boldsymbol{\theta}}_n = (\widetilde{\alpha}_n, \widetilde{\beta}_n)^\top$ de $\boldsymbol{\theta}_0 = (\alpha_0, \beta_0)^\top$ é tal que $\widetilde{\boldsymbol{\theta}}_n \xrightarrow{\text{P}} \boldsymbol{\theta}_0$. Além disso, $\widetilde{\boldsymbol{\theta}}_n$ é assintoticamente normal com média $\boldsymbol{\theta}_0$ e matriz de covariâncias dada por*

$$\boldsymbol{\Sigma}_n = \frac{\nu+3}{\nu+1}\frac{1}{(S_{xx}^0)^2}\begin{pmatrix} \Delta(S_{xx}^0)^2 + K\overline{x}_0^2 & -K\overline{x}_0 \\ -K\overline{x}_0 & K \end{pmatrix}, \tag{3.58}$$

*onde $\Delta = 1 + \beta_0^2$, $K = \Delta S_{xx}^0 + A_f$ e $A_f$ definido em (3.55).*

**Prova:** Primeiramente provaremos que as condições de "regularidade" A1 – A5 e A7 são satisfeitas neste modelo funcional $t$-Student. Com efeito,

A3. De (3.49) segue que

$$|W_f(u)| \leq \frac{3}{2} \quad \text{e} \quad |W_f'(u)| \leq \frac{5}{2},$$

onde $W_f'(u) = f''(u)/f'(u) = -\frac{\nu+4}{2}(\nu+u)^{-1}$, $u \geq 0$.

A4. Da relação (3.51), temos que

$$|h_i(\mathbf{Z}_i; \boldsymbol{\theta})| \leq \log 2\pi + \frac{\nu+2}{2}\log(1+t_i^2).$$

Desta desigualdade e da demonstração do Teorema 3.1 segue que, para $\boldsymbol{\theta} \in B(\boldsymbol{\theta}_0, \delta)$, $\delta > 0$,

$$|h_i(\mathbf{Z}_i; \boldsymbol{\theta})| \leq \log 2\pi + \frac{\nu+2}{2} \log(1 + d_0^2(\mathbf{Z}_i)) = d_i(\mathbf{Z}_i),$$

onde $d_0(\mathbf{Z}_i) = \delta(1 + |x_{i0}|) + \|\mathbf{Z}_i - \mathbf{a}_0 - \mathbf{b}_0 x_{i0}\|$, com $\mathbf{Z}_i - \mathbf{a}_0 - \mathbf{b}_0 x_{i0} \overset{\text{iid}}{\sim} t_2(0, I_2; \nu)$, $\nu > 4$. Além disso,

$$\begin{aligned} d_i^2(\mathbf{Z}_i) &\leq 4\left\{(\log 2\pi)^2 + \frac{(\nu+2)^2}{4}\left[\log(1 + d_0^2(\mathbf{Z}_i))\right]\right\} \\ &= 4(\log 2\pi)^2 + (\nu+2)^2\left[\log(1 + d_0^2(\mathbf{Z}_i)\right]^2 \end{aligned}$$

e $e^u \geq 1 + u$, para todo $u \geq 0$, o que implica que

$$E_0\left[\left\{\log(1 + d_0^2(\mathbf{Z}_i)\right\}^2\right] \leq E_0[d_0^4(\mathbf{Z}_i)] \leq 2^4\left\{\delta^4(1 + |x_{i0}|)^4 + E_0\left[\|\mathbf{Z}_i - \mathbf{a}_0 - \mathbf{b}_0 x_{i0}\|^4\right]\right\}$$

e, conseqüentemente, por A6 temos que

$$\limsup \frac{1}{n} \sum_{i=1}^{n} E_0[d_i^2(\mathbf{Z}_i)] < \infty.$$

Por outro lado, para $\boldsymbol{\theta} \in B(\boldsymbol{\theta}_0, \delta)$, $\delta > 0$, tem-se que

$$|q_{i\theta_k\theta_j}(\mathbf{Z}_i; \boldsymbol{\theta})| \leq d_{ikj}(\mathbf{Z}_i), \quad i = 1, \ldots, n, \tag{3.59}$$

onde $\boldsymbol{\theta} = (\theta_1, \theta_2) = (\alpha, \beta)^\top$ e

$$d_{ikj}(\mathbf{Z}_i) = \begin{cases} 2(\nu+2) + 3 & ; \theta_k = \theta_j = \alpha, \\ 2(\nu+2)(|x_{i0}| + 1 + |s_i|) & ; \theta_k = \alpha,\ \theta_j = \beta, \\ 4(\nu+2)(|x_{i0}| + 1 + |s_i|)^2 & ; \theta_k = \theta_j = \beta. \end{cases}$$

Para $\boldsymbol{\theta} \in B(\boldsymbol{\theta}_0, \delta)$

$$|s_i| \leq d_0(\mathbf{Z}_i), \tag{3.60}$$

onde

$$E_0[d_0^4(\mathbf{Z}_i)] \leq 2^4\{\delta^4(1 + |x_{i0}|)^4 + E_0[\|\mathbf{Z}_i - \mathbf{a}_0 - \mathbf{b}_0 x_{i0}\|^4], \tag{3.61}$$

com $\mathbf{Z}_i - \mathbf{a}_0 - \mathbf{b}_0 x_{i0} \overset{\text{iid}}{\sim} t_2(\mathbf{0}, \mathbf{I}_2; \nu)$, $\nu > 4$. Assim, de (3.59), (3.60), (3.61) juntamente com a condição A6, segue que

$$\limsup \frac{1}{n} \sum_{i=1}^{n} E_0[d_{ikj}^2(\mathbf{Z}_i)] < \infty,$$

de onde segue que a condição A4 é satisfeita.

A5. De (3.30) e (3.49), temos que

$$|q_{i\alpha}(\mathbf{Z}_i;\boldsymbol{\theta}_0)| \leq \frac{\nu+2}{2}$$

e

$$|q_{i\beta}(\mathbf{Z}_i;\boldsymbol{\theta}_0)| \leq \frac{\nu+2}{2}(1+|x_{i0}|+|s_{i0}|)$$

Logo, para $\gamma = 2$, temos que

$$E_0[|q_{i\alpha}(\mathbf{Z}_i;\boldsymbol{\theta}_0)|^{2+2}] \leq \frac{\nu+2}{2}$$

e

$$E_0[|q_{i\beta}(\mathbf{Z}_i;\boldsymbol{\theta}_0)|^{2+2}] \leq (\nu+2)^4\{(1+|x_{i0}|)^4 + E_0[s^4]\}$$

Portanto, da condição A6, segue que a condição A5 é satisfeita.

A7. Como a terceira derivada da função $f(u)$, $u \geq 0$, existe, então será suficiente provar a condição A7'. Depois de algumas extensas manipulações algébricas, tem-se que

$$\begin{aligned}
& |q_{i\alpha\alpha}(\mathbf{Z}_i;\boldsymbol{\theta})| \leq 5(\nu+2), \\
& |q_{i\alpha\alpha\beta}(\mathbf{Z}_i;\boldsymbol{\theta})| \leq 7(\nu+2)(|x_{i0}|+1+|s_i|)+3, \\
& |q_{i\beta\beta\alpha}(\mathbf{Z}_i;\boldsymbol{\theta})| \leq 5(\nu+2)(|x_{i0}|+1+|s_i|)^2 + 2(\nu+2)(|x_{i0}+1+|s_i|) \\
\text{e} \quad & |q_{i\beta\beta\beta}(\mathbf{Z}_i;\boldsymbol{\theta})| \leq 3(\nu+2)(|x_{i0}|+1+|s_i|)^3 + 3(\nu+2)(|x_{i0}+1+|s_i|).
\end{aligned}$$

Logo, para $\boldsymbol{\theta} \in B(\boldsymbol{\theta}_0, \delta)$, $\delta > 0$, e a condição A6 satisfeita, tem-se que

$$|s_i| \leq d_0(\mathbf{Z}_i) \leq \|\mathbf{Z}_i - \mathbf{a}_0 - \mathbf{b}_0 x_{i0}\| + \delta(1+M),$$

com $E_0[d_0^3(\mathbf{Z}_i)]$ finito. Portanto, dos resultados acima temos que existe uma função positiva $d_{ikjl}(\mathbf{Z}_i)$ tal que

$$|q_{i\theta_k\theta_j\theta_l}(\mathbf{Z}_i;\boldsymbol{\theta})| \leq d_{ikjl}(\mathbf{Z}_i),$$

para $\boldsymbol{\theta} \in B(\boldsymbol{\theta}_0, \delta)$, $\delta > 0$ e

$$\limsup \frac{1}{n} \sum_{i=1}^{n} E_0[d_{ikjl}(\mathbf{Z}_i)] < \infty,$$

com $\boldsymbol{\theta} = (\theta_1, \theta_2)^\top = (\alpha, \beta)^\top$ e $k, j, l = 1, 2$, o que mostra que a condição A7' é satisfeita. Portanto, temos que as condições A1 – A7 são satisfeitas para o modelo funcional $t$-Student.

A demonstração do Teorema 4.1 segue então do Teorema 3.2 juntamente com (3.56) e (3.57) □

**Corolário 3.2** *Sob as condições do Teorema 3.4 e a condição A8, tem-se que*

$$e_{\widehat{\beta}_{GLS}, \widetilde{\beta}_n} = \frac{(\nu-2)(\nu+3)}{\nu(\nu+1)} \left( \frac{\Delta\sigma^*_{xx} + \frac{\nu}{\nu-4}}{\Delta\sigma^*_{xx} + \frac{\nu+3}{\nu-1}} \right), \quad \nu > 4, \tag{3.62}$$

*o que implica que $\widetilde{\beta}_n$ é mais eficiente que $\widehat{\beta}_{GLS}$.*

**Prova:** A demonstração de (3.62) segue do Corolário 3.1 juntamente com (3.55) e do fato que no modelo $t$-Student $\kappa + 1 = \frac{\nu-2}{\nu-4}$ e $\delta = \frac{\nu}{\nu-2}$. Além disso, da desigualdade

$$\frac{(\nu-2)(\nu+3)}{\nu(\nu+1)} \leq 1 \quad \text{e} \quad \frac{\nu+3}{\nu-1} \leq \frac{\nu}{\nu-4}$$

segue que $e_{\widehat{\beta}_{GLS}\widetilde{\beta}_n} < 1$, $\nu > 4$, e, conseqüentemente, o estimador de máxima verossimilhança $\widetilde{\beta}_n$ é assintoticamente mais eficiente que o estimador $\widehat{\beta}_{GLS}$ para $\nu > 4$.

## 3.4 Inferência no modelo funcional elíptico multi-univariado

Nesta seção, estudaremos as propriedades assintóticas dos EMV dos parâmetros estruturais do modelo funcional elíptico em (3.5) para $q > 1$. Similaramente ao ocorrido no modelo funcional univariado, denotaremos por $\boldsymbol{\theta}_0 = (\boldsymbol{\alpha}_0^\top, \boldsymbol{\beta}_0^\top)^\top$ e $x_{10}, \ldots, x_{n0}$ os verdadeiros valores de $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top)^\top$ e $x_1, \ldots, x_n$, respectivamente, e as esperanças serão tomadas considerando estes valores.

De (3.6) temos que, para uma amostra de tamanho $n$, a função de log-verossimilhança é dada por

$$\sum_{i=1}^{n} \log f_i(\mathbf{Z}_i; \boldsymbol{\theta}, x_i) = \sum_{i=1}^{n} \log f(\|\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_i\|^2), \tag{3.63}$$

onde $\mathbf{Z}_i = (\mathbf{Y}_i^\top, X_i)^\top$, $\mathbf{a} = (\boldsymbol{\alpha}^\top, 0)^\top$ e $\mathbf{b} = (\boldsymbol{\beta}^\top, 1)^\top$ são vetores $p = q + 1$-dimensionais (veja a Seção 3.1) e $f(\mathbf{t}^\top \mathbf{t})$, $\mathbf{t} \in \mathbb{R}^p$, sendo uma densidade esférica $p$-dimensional que satisfaz as seguintes condições:

B1. $f \in C^{(2)}$ e é decrescente em $(0, +\infty)$;

B2. $r^{p+3} f(r^2)$ é integrável sobre $[0, +\infty)$, isto é, a distribuição de $\epsilon_1$ tem quartos momentos finitos.

Sejam

$$\mathbf{b}_1 = \frac{1}{\|\mathbf{b}\|}\mathbf{b}, \quad \mathbf{B}_1 = \mathbf{b}_1 \mathbf{b}_1^\top \quad \text{e} \quad \mathbf{B}_2 = \mathbf{I}_p - \mathbf{B}_1 \tag{3.64}$$

e os vetores aleatórios

$$\mathbf{R}_i = \mathbf{B}_1(\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_{i0}), \quad \mathbf{T}_i = \mathbf{B}_2(\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_{i0}) \quad \text{e} \quad r_i = \mathbf{b}_1^\top \mathbf{R}_i. \tag{3.65}$$

Notemos que $\mathbf{R}_i = R_i(\boldsymbol{\theta})$ e $\mathbf{T}_i = T_i(\boldsymbol{\theta})$ são ortogonais para cada $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, pois $\mathbf{B}_1\mathbf{B}_2 = \mathbf{B}_2\mathbf{B}_1 = 0$. Além disso, a distribuição conjunta destes vetores é elíptica singular. Para mostrar esta afirmação, precisamos do seguinte resultado.

**Lema 3.5** *Seja* $\mathbf{X} = (\mathbf{X}_1^\top, \mathbf{X}_2^\top)^\top \sim E\ell_p(\mathbf{0}, \mathbf{I}_p)$, *onde* $\mathbf{X}_i$ *é* $p_i$*-dimensional,* $i = 1, 2$ *e* $p_1 + p_2 = p$. *Seja* $\mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top$ *um vetor aleatório* $2p$*-dimensional, definido por* $\mathbf{Y}_i = \mathbf{A}_i\mathbf{X}_i$, $i = 1, 2$, *onde* $\mathbf{A}_1$ *e* $\mathbf{A}_2$ *são matrizes simétricas, ortogonais idempotentes, tais que* $\mathbf{A}_1 + \mathbf{A}_2 = I_p$ *e posto*$(\mathbf{A}_i) = p_i$, $i = 1, 2$. *Então,*

*i)* $\mathbf{Y} \sim E\ell_{2p}(\mathbf{0}, \mathbf{A})$, *onde* $\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & 0 \\ 0 & \mathbf{A}_2 \end{pmatrix}$;

*ii) Existe* $\mathbf{\Gamma} \in O_p$ *tal que*

$$\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} \stackrel{d}{=} \begin{pmatrix} \mathbf{\Gamma}^\top \mathbf{Q}_2^\top \mathbf{X}_2 \\ \mathbf{\Gamma}^\top \mathbf{Q}_1^\top \mathbf{X}_1 \end{pmatrix},$$

*onde* $\mathbf{Q}_1 = [\mathbf{I}_{p_1}, 0]$ *e* $\mathbf{Q}_2 = [0, \mathbf{I}_{p_2}]$ *são matrizes de dimensão* $p_1 \times p$ *e* $p_2 \times p$*, respectivamente.*

**Prova:** i) Seja $\mathbf{C}$ a matriz de dimensão $2p \times 2p$ definida por

$$\mathbf{C} = \begin{pmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{pmatrix}.$$

Assim, pelo Teorema A.3 do apêndice, temos que $\mathbf{Y} = \mathbf{CX} \sim E\ell_{2p}(\mathbf{0}, \mathbf{CC}^\top)$, onde $\mathbf{CC}^\top = \mathbf{A}$, pois $\mathbf{A}_i^2 = \mathbf{A}_i$, $i = 1, 2$ e $\mathbf{A}_1\mathbf{A}_2 = \mathbf{A}_2\mathbf{A}_1 = \mathbf{0}$.

ii) Como $\mathbf{A}_1$ e $\mathbf{A}_2 = \mathbf{I}_p - \mathbf{A}_1$ são matrizes idempotentes, existe $\mathbf{\Gamma} \in O_p$ tal que $\mathbf{\Gamma A}_i \mathbf{\Gamma}^\top = \mathbf{Q}_i^\top \mathbf{Q}_i$, ou seja, $\mathbf{A}_i = \mathbf{\Gamma}^\top \mathbf{Q}_i^\top \mathbf{Q}_i \mathbf{\Gamma}$, $i = 1, 2$, onde $\mathbf{Q}_1 = [\mathbf{I}_{p_1}\ 0]$ e $\mathbf{Q}_2 = [0\ \mathbf{I}_{p_2}]$. Como $\mathbf{\Gamma} \in O_p$ e $\mathbf{X} \sim E\ell_p(\mathbf{0}, \mathbf{I}_p)$ satisfazendo a relação acima, então temos que $\mathbf{X} \stackrel{d}{=} \mathbf{\Gamma X}$, que implica que

$$\begin{aligned}
\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} &= \begin{pmatrix} \mathbf{A}_1\mathbf{X} \\ \mathbf{A}_2\mathbf{X} \end{pmatrix} = \begin{pmatrix} \mathbf{\Gamma}^\top \mathbf{Q}_2^\top \mathbf{Q}_2 \mathbf{\Gamma X} \\ \mathbf{\Gamma}^\top \mathbf{Q}_1^\top \mathbf{Q}_1 \mathbf{\Gamma X} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{\Gamma}^\top \mathbf{Q}_2^\top \mathbf{Q}_2 \\ \mathbf{\Gamma}^\top \mathbf{Q}_1^\top \mathbf{Q}_1 \end{pmatrix} (\mathbf{\Gamma X}) \stackrel{d}{=} \begin{pmatrix} \mathbf{\Gamma}^\top \mathbf{Q}_2^\top \mathbf{Q}_2 \\ \mathbf{\Gamma}^\top \mathbf{Q}_1^\top \mathbf{Q}_1 \end{pmatrix} \mathbf{X} \\
&= \begin{pmatrix} \mathbf{\Gamma}^\top \mathbf{Q}_2^\top \mathbf{Q}_2 \mathbf{X} \\ \mathbf{\Gamma}^\top \mathbf{Q}_1^\top \mathbf{Q}_1 \mathbf{X} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{\Gamma}^\top \mathbf{Q}_2^\top \mathbf{X}_2 \\ \mathbf{\Gamma}^\top \mathbf{Q}_1^\top \mathbf{X}_1 \end{pmatrix}.
\end{aligned}$$

□

O seguinte resultado é uma conseqüência imediata do lema anterior.

**Corolário 3.3** *Para qualquer função contínua* $g : \mathbb{R}^{2p} \to \mathbb{R}$*, tem-se que*

$$g\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} \stackrel{d}{=} g\begin{pmatrix} \mathbf{\Gamma}^\top \mathbf{Q}_2^\top \mathbf{X}_2 \\ \mathbf{\Gamma}^\top \mathbf{Q}_1 \mathbf{X}_1 \end{pmatrix}.$$

**Corolário 3.4** *Para cada* $i$*,*

*i)* $(\mathbf{T}_i^\top, \mathbf{R}_i^\top)^\top \sim E\ell_{2p}(\mathbf{C}[(\mathrm{a}_0 - \mathrm{a}) + (\mathrm{b}_0 - \mathrm{b})x_{i0}], \mathbf{C}\mathbf{C}^\top)$, *onde* $\mathbf{C} = \binom{\mathbf{B}_2}{\mathbf{B}_1}$ *e* $\mathbf{C}\mathbf{C}^\top = \begin{pmatrix} \mathbf{B}_2 & 0 \\ 0 & \mathbf{B}_1 \end{pmatrix}$;

*ii)* $(\mathbf{T}_{i0}^\top, \mathbf{R}_{i0}^\top)^\top = (\mathbf{T}_i^\top(\boldsymbol{\theta}_0), \mathbf{R}_i^\top(\boldsymbol{\theta}_0))^\top \sim E\ell_{2p}(\mathbf{0}, \mathbf{C}_0\mathbf{C}_0^\top)$ *e existe* $\boldsymbol{\Gamma} \in O_p$ *tal que*

$$\begin{pmatrix} \mathbf{T}_{i0} \\ \mathbf{R}_{i0} \end{pmatrix} \stackrel{d}{=} \begin{pmatrix} \boldsymbol{\Gamma}^\top \mathbf{Q}_1 \mathbf{e}_1 \\ \boldsymbol{\Gamma}^\top \mathbf{Q}_2^\top u_1 \end{pmatrix}, \tag{3.66}$$

*onde* $\mathbf{Q}_1 = [\mathbf{I}_q\ 0]$ *tem dimensão* $q \times p$, $\mathbf{Q}_2 = [\mathbf{0}\ 1]$ *tem dimensão* $1 \times p$ *e* $(\mathbf{e}_1^\top, u_1)^\top \sim E\ell_p(\mathbf{0}, \mathbf{I}_p; f)$.

**Prova:** A prova de i) é análoga à demonstração do Lema 4.1 (i).

A primeira parte da demonstração de ii) segue diretamente da parte i) fazendo $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Para a segunda parte, usamos o Lema 4.1 (ii), com $\mathbf{A}_1 = \mathbf{B}_2$, $\mathbf{A}_2 = \mathbf{B}_1$ e $\boldsymbol{\epsilon}_1 = (\mathbf{e}_1^\top, u_1)^\top \stackrel{d}{=} \mathbf{Z}_i - \mathbf{a}_0 - \mathbf{b}_0 x_{i0} \sim E\ell_p(\mathbf{0}, \mathbf{I}_p)$, onde $\mathbf{Q}_1 = [\mathbf{I}_1, 0] : q \times p$ e $\mathbf{Q}_2 = [\mathbf{0}, 1] : 1 \times p$. □

### 3.4.1 Estimação por máxima verossimilhança

Nesta seção, como no modelo funcional elíptico univariado, estudaremos o comportamento assintótico do estimador de máxima verossimilhança dos parâmetros estruturais do modelo estatístico definido em (3.6), com $q > 1$. Primeiramente, vamos determinar o EMV condicional de $x_i$, $i = 1, \ldots, n$, dado $\boldsymbol{\theta}$.

**Lema 3.6** *Considere o modelo definido em (3.5) satisfazendo a condição B1. Então, o estimador de máxima verossimilhança condicional de* $x_i$ *dado* $\theta$, $i = 1, \ldots, n$, *é dado por*

$$\widehat{x}_i = \frac{\mathbf{b}^\top}{\|\mathbf{b}\|^2}(\mathbf{Z}_i - \mathbf{a}), \tag{3.67}$$

*e*

$$h_i(\mathbf{Z}_i; \boldsymbol{\theta}) = \log f(\|\mathbf{T}_i\|^2), \tag{3.68}$$

*onde* $\mathbf{T}_i$ *e* $h_i(\mathbf{Z}_i; \boldsymbol{\theta})$ *são como em (3.65) e (3.7), respectivamente.*

**Prova:** A demonstração de (3.67) e (3.68) é análoga ao caso univariado (veja Lema 3.2).

□

Considerando $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top)^\top = (\alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_q)^\top \in \boldsymbol{\Theta}$, $W_f(u) = f'(u)/f(u)$ e $\Delta_f(u) = W_{f'}(u)W_f(u) - W_f^2(u)$, $u \geq 0$, de (3.9), (3.10), (3.11) e (3.68), segue que

$$q_{i\theta_k}(\mathbf{Z}_i; \boldsymbol{\theta}) = W_f(\|\mathbf{T}_i\|^2)\left(\frac{\partial}{\partial\theta_k}\|\mathbf{T}_i\|^2\right), \quad k = 1, \ldots, 2q, \tag{3.69}$$

$$q_{i\theta_k\theta_j}(\mathbf{Z}_i; \boldsymbol{\theta}) = \Delta_f(\|\mathbf{T}_i\|^2)\left(\frac{\partial}{\partial\theta_k}\|\mathbf{T}_i\|^2\right)\left(\frac{\partial}{\partial\theta_j}\|\mathbf{T}_i\|^2\right) + W_f(\|\mathbf{T}_i\|^2)\left(\frac{\partial}{\partial\theta_k\theta_j}\|\mathbf{T}_i\|^2\right), \tag{3.70}$$

$$q_{i\theta_k,\theta_j}(\mathbf{Z}_i; \boldsymbol{\theta}) = W_f^2(\|\mathbf{T}_i\|^2)\left(\frac{\partial}{\partial\theta_k}\|\mathbf{T}_i\|^2\right)\left(\frac{\partial}{\partial\theta_j}\|\mathbf{T}_i\|^2\right), \tag{3.71}$$

$k, j = 1, \ldots, 2q$, onde $\theta_l = \alpha_l$, $l = 1, \ldots, q$ e $\theta_{l+q} = \beta_l$, $l = 1, \ldots, q$. Além disso, após algumas manipulações algébricas, mostra-se que

$$\frac{\partial\|\mathbf{T}_i\|^2}{\partial\alpha_j} = -2(\mathbf{d}_j^\top\mathbf{T}_i), \qquad \frac{\partial\|\mathbf{T}_i\|^2}{\partial\beta_j} = -2(\mathbf{d}_j^\top\mathbf{T}_i)x_{i0} - \frac{2}{\|\mathbf{b}\|}(\mathbf{d}_j^\top\mathbf{T}_i)r_i,$$

$$\frac{\partial^2\|\mathbf{T}_i\|^2}{\partial\alpha\partial\beta} = 2(\mathbf{d}_j^\top\mathbf{B}_2\mathbf{d}_k),$$

$$\frac{\partial^2\|\mathbf{T}_i\|^2}{\partial\alpha_j\partial\beta_k} = \frac{\partial^2\|\mathbf{T}_i\|^2}{\partial\alpha_j\partial\alpha_k}x_{i0} + \frac{2}{\|\mathbf{b}\|}(\mathbf{d}_j^\top\mathbf{B}_2\mathbf{d}_k)r_i + \frac{2}{\|\mathbf{b}\|}(\mathbf{d}_j^\top\mathbf{b}_1)(\mathbf{d}_k^\top\mathbf{T}_i),$$

$$\frac{\partial^2\|\mathbf{T}_i\|^2}{\partial\beta_j\partial\beta_k} = \frac{\partial^2\|\mathbf{T}_i\|^2}{\partial\alpha_j\partial\alpha_k}x_{i0}^2 + \frac{4x_{i0}}{\|\mathbf{b}\|}(\mathbf{d}_j^\top\mathbf{B}_2\mathbf{d}_k)r_i + \frac{2x_{i0}}{\|\mathbf{b}\|}\mathbf{b}_1^\top\mathbf{D}_{(j,k)}\mathbf{T}_i$$

$$+\frac{2}{\|\mathbf{b}\|^2}\mathbf{b}_1^\top\mathbf{D}_{(j,k)}\mathbf{T}_i\Gamma_i - \frac{2}{\|\mathbf{b}\|^2}(\mathbf{T}_i^\top\mathbf{D}_{jk}\mathbf{T}_i) + \frac{2}{\|\mathbf{b}\|^2}(\mathbf{d}_j^\top\mathbf{B}_2\mathbf{d}_k)r_i^2,$$

com $\mathbf{D}_{jk} = \mathbf{d}_j\mathbf{d}_k^\top$, $\mathbf{D}_{(j,k)} = \mathbf{D}_{jk} + \mathbf{D}_{kj}$ e $\mathbf{d}_j = (\boldsymbol{\delta}_j^\top, 0)^\top \in \mathbb{R}^p$, $p = q+1$, com $\boldsymbol{\delta}_j = (0, \ldots, 1, \ldots, 0)^\top \in \mathbb{R}^q$, $j = 1, \ldots, q$, com 1 na $j$-ésima posição. □

O seguinte resultado será de bastante utilidade para determinar a matriz de covariâncias assintóticas do EMV do parâmetro estrutural do modelo funcional que estamos considerando.

**Lema 3.7** *Considere o modelo definido por (3.5) satisfazendo as condições B1 e B2. Então, as matrizes* $E_0[A_n(\boldsymbol{\theta}_0)]$ *e* $V_n(\boldsymbol{\theta}_0)$ *definidas em (3.12) e (3.13), respectivamente, são dadas por*

$$E_0[A_n(\boldsymbol{\theta}_0)] = 2C_f \begin{pmatrix} 1 & \overline{x}_0 \\ \overline{x}_0 & S_x^0 + \frac{1}{\|\mathbf{b}_0\|^2}\frac{B_f}{C_f} \end{pmatrix} \otimes \mathbf{B}_{30}, \tag{3.72}$$

$$V_n(\boldsymbol{\theta}_0) = \frac{4}{q} E_0[W_f^2(\|\mathbf{e}_1\|^2)\|\mathbf{e}_1\|^2] \begin{pmatrix} 1 & \overline{x}_0 \\ \overline{x}_0 & S_x^0 + \frac{1}{\|\mathbf{b}_0\|^2} A_f \end{pmatrix} \otimes \mathbf{B}_{30}, \tag{3.73}$$

*onde*

$$\overline{x}_0 = \frac{1}{n}\sum_{i=1}^n x_{i0}, \quad S_x^0 = \frac{1}{n}\sum_{i=1}^n x_{i0}^2, \quad \mathbf{B}_{30} = I_q - \frac{\boldsymbol{\beta}_0\boldsymbol{\beta}_0^\top}{\|\mathbf{b}_0\|^2},$$

$$A_f = \frac{E_0[W_f^2(\|\mathbf{e}_1\|^2)a(\|\mathbf{e}_1\|^2)\|\mathbf{e}_1\|^2]}{E_0[W_f^2(\|\mathbf{e}_1\|^2)\|\mathbf{e}_1\|^2]},$$

$$B_f = E_0[W_f(\|\mathbf{e}_1\|^2)\{a(\|\mathbf{e}_1\|^2) - \frac{1}{q}\|\mathbf{e}_1\|^2\}] + \frac{2}{q}E_0[\Delta_f(\|\mathbf{e}_1\|^2)a(\|\mathbf{e}_1\|^2)\|\mathbf{e}_1\|^2]$$

*e*

$$C_f = E_0[W_f(\|\mathbf{e}_1\|^2)] + \frac{2}{q}E_0[\Delta_f(\|\mathbf{e}_1\|^2)\|\mathbf{e}_1\|^2],$$

*com* $a(\|\mathbf{e}_1\|^2) = E_0[u_1^2|\mathbf{e}_1]$ *e* $(\mathbf{e}_1^\top, u_1)^\top \sim E\ell_p(\mathbf{0}, \mathbf{I}_p; f)$.

**Prova:** De (3.70) e usando os fatos dados no Corolário 4.2, segue que

$$\begin{pmatrix} r_{i0} \\ \mathbf{T}_{i0} \end{pmatrix} \overset{d}{=} \begin{pmatrix} \mathbf{b}_{10}^\top \boldsymbol{\Gamma} \mathbf{Q}_2^\top u_1 \\ \boldsymbol{\Gamma}^\top \mathbf{Q}_1^\top \mathbf{e}_1 \end{pmatrix}, \tag{3.74}$$

onde $\boldsymbol{\Gamma}$, $\mathbf{Q}_1$ e $\mathbf{Q}_2$ são como em (3.66) e

$$E_0[q_{i\alpha_k\alpha_j}(\mathbf{Z}_i;\boldsymbol{\theta}_0)] = 2\{2E_0[\Delta_f(\|\mathbf{T}_{i0}\|^2)(\mathbf{T}_{i0}^\top \mathbf{D}_{kj}\mathbf{T}_{i0})] + (\mathbf{d}_j^\top \mathbf{B}_2 \mathbf{d}_k)E_0[W_f(\|\mathbf{T}_{i0}\|^2)]\}. \tag{3.75}$$

Além disso, como as funções $W_f$ e $\Delta_f$ são contínuas, então o Corolário 4.1 implica que

$$\Delta_f(\|\mathbf{T}_{i0}\|^2)(\mathbf{T}_{i0}^\top \mathbf{D}_{jk}\mathbf{T}_{i0}) \stackrel{d}{=} \Delta_f(\|\mathbf{e}_1\|^2)(\mathbf{e}_1^\top \mathbf{Q}_1 \boldsymbol{\Gamma}\mathbf{D}_{kj}\boldsymbol{\Gamma}\mathbf{Q}_1^\top \mathbf{e}_1), \quad \text{e} \tag{3.76}$$

$$W_f(\|\mathbf{T}_{i0}\|^2) \stackrel{d}{=} W_f(\|\mathbf{e}_1\|^2). \tag{3.77}$$

Como $\|\mathbf{e}_1\|$ é independente de $\mathbf{u}^{(q)} \stackrel{d}{=} \mathbf{e}_1/\|\mathbf{e}_1\|$, então, usando os Teoremas A.1 e A.13 do apêndice, temos que

$$\begin{aligned} E_0[W_f(\|\mathbf{T}_{i0}\|^2)(\mathbf{T}_{i0}^\top \mathbf{D}_{kj}\mathbf{T}_{i0})] &= E_0[\Delta_f(\|\mathbf{e}_1\|^2)(\mathbf{e}_1^\top \mathbf{Q}_1\boldsymbol{\Gamma}\mathbf{D}_{kj}\boldsymbol{\Gamma}^\top \mathbf{Q}_1^\top \mathbf{e}_1] \\ &= \frac{1}{q}\mathrm{tr}(\mathbf{Q}_1\boldsymbol{\Gamma}\mathbf{D}_{kj}\boldsymbol{\Gamma}^\top\mathbf{Q}_1^\top)E_0[\Delta_f(\|\mathbf{e}_1\|^2)\|\mathbf{e}_1\|^2] \\ &= \frac{1}{q}(\mathbf{d}_j^\top \mathbf{B}_2\mathbf{d}_k)E_0[\Delta_f(\|\mathbf{e}_1\|^2)\|\mathbf{e}_1\|^2]. \end{aligned} \tag{3.78}$$

Assim, de (3.75), (3.76), (3.77) e (3.78), temos que

$$\begin{aligned} E_0[q_{i\alpha_k\alpha_j}(\mathbf{Z}_i;\boldsymbol{\theta}_0)] &= 2(\mathbf{d}_j^\top \mathbf{B}_{20}\mathbf{d}_k)\{E_0[W_f(\|\mathbf{e}_1\|^2)] + \frac{2}{q}E_0[\Delta_f(\|\mathbf{e}_1\|^2)\|\mathbf{e}_1\|^2]\} \\ &= 2b_{jk}\{E_0[W_f(\|\mathbf{e}_1\|^2) + \frac{2}{q}E_0[\Delta_f(\|\mathbf{e}_1\|^2)\|\mathbf{e}_1\|^2]\}, \end{aligned}$$

onde $b_{jk}^0 = \mathbf{d}_j^\top \mathbf{B}_{20}\mathbf{d}_k$ é o elemento $(j,k)$ da matriz

$$\mathbf{B}_{30} = \mathbf{I}_q - \frac{\boldsymbol{\beta}_0\boldsymbol{\beta}_0^\top}{\|\mathbf{b}_0\|^2},$$

pois $\mathbf{d}_j^\top \mathbf{B}_{20}\mathbf{d}_k = \boldsymbol{\delta}_j^\top \mathbf{B}_{30}\boldsymbol{\delta}_k$, com $\mathbf{d}_j = (\boldsymbol{\delta}_j^\top, 0)^\top$ e $\boldsymbol{\delta}_j = (0,\ldots,1,\ldots,0)^\top \in \mathbb{R}^q$, 1 na $j$-ésima posição.

Similarmente, usando o fato que $E_0[r_{i0}|\mathbf{T}_{i0}] = 0$, já que $(r_{i0}, \mathbf{T}_{i0}^\top)^\top \stackrel{\text{iid}}{\sim} E\ell_{p+1}(\mathbf{0}, \mathbf{B}_0)$, com $\mathbf{B}_0 = \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{B}_{20} \end{pmatrix}$, mostramos que

$$E_0[q_{i\alpha_k\beta_j}(\mathbf{Z}_i, \boldsymbol{\theta}_0)] = x_{i0}E_0[q_{i\alpha_k\alpha_j}(\mathbf{Z}_i, \boldsymbol{\theta}_0)].$$

Agora, de (3.70) e do fato que $E_0[r_{i0}|\mathbf{T}_{i0}] = 0$, segue que

$$\begin{aligned} &E_0[q_{i\beta_k\beta_j}(\mathbf{Z}_i;\boldsymbol{\theta}_0)] = x_{i0}^2 E_0[q_{i\alpha_k\alpha_j}(\mathbf{Z}_j;\boldsymbol{\theta}_0)] \\ &+\frac{2}{\|\mathbf{b}_0\|^2}\{2E_0[\Delta_f(\|\mathbf{T}_{i0}\|^2)(\mathbf{T}_{i0}^\top \mathbf{D}_{jk}\mathbf{T}_{i0})r_{i0}^2] + (\mathbf{d}_j^\top \mathbf{B}_{20}\mathbf{d}_k)E_0[W_f(\|\mathbf{T}_{i0}\|^2)r_{i0}^2] \\ &-E_0[W_f(\|\mathbf{T}_{i0}\|^2)(\mathbf{T}_{i0}^\top \mathbf{D}_{jk}\mathbf{T}_{i0})]\}. \end{aligned}$$

De (3.74) e considerando que $W_f$ e $\Delta_f$ são funções contínuas, segue do Corolário 4.1 que

$$\begin{aligned}
\Delta_f(\|\mathbf{T}_{i0}\|^2)(\mathbf{T}_{i0}^\top \mathbf{D}_{jk}\mathbf{T}_{i0})r_{i0}^2 &\stackrel{\mathrm{d}}{=} \Delta_f(\|\mathbf{e}_1\|^2)(\mathbf{e}_1^\top \mathbf{Q}_1 \boldsymbol{\Gamma} \mathbf{D}_{jk} \boldsymbol{\Gamma}^\top \mathbf{Q}_1^\top \mathbf{e}_1)u_1^2 \\
W_f(\|\mathbf{T}_{i0}\|^2)r_{i0}^2 &\stackrel{\mathrm{d}}{=} W_f(\|\mathbf{e}_1\|^2)u_1^2
\end{aligned}$$

e

$$W_f(\|\mathbf{T}_{i0}\|^2)(\mathbf{T}_{i0}^\top \mathbf{D}_{jk}\mathbf{T}_{i0}) \stackrel{\mathrm{d}}{=} W_f(\|\mathbf{e}_1\|^2)(\mathbf{e}_1^\top \mathbf{Q}_1 \boldsymbol{\Gamma} \mathbf{D}_{jk} \boldsymbol{\Gamma}^\top \mathbf{Q}_1^\top \mathbf{e}_1).$$

Considerando estes últimos resultados, temos que

$$E_0[q_{i\beta_k\beta_j}(\mathbf{Z}_i;\boldsymbol{\theta}_0)] = x_{i0}E_0[q_{i\alpha_k\alpha_j}(\mathbf{Z}_i;\boldsymbol{\theta}_0)] + \frac{2}{\|\mathbf{b}_0\|^2} b_{jk}^0 \mathbf{B}_f.$$

Portanto, a matriz $E_0[A_n(\boldsymbol{\theta}_0)]$ segue de (3.12).

Similarmente, de (3.71), (3.74) e da simetria da distribuição de $(r_{i0}, \mathbf{T}_{i0}^\top)^\top$ em torno da origem, segue que as entradas da matriz $V_n(\boldsymbol{\theta}_0)$ são dadas por

$$\begin{aligned}
E_0[q_{i\alpha_j,\alpha_k}(\mathbf{Z}_i;\boldsymbol{\theta}_0)] &= \frac{4}{q} b_{jk}^0 E_0[W_f^2(\|\mathbf{e}_1\|^2)\|\mathbf{e}_1\|^2] \\
E_0[q_{i\alpha_j,\beta_k}(\mathbf{Z}_i;\boldsymbol{\theta}_0)] &= x_{i0}E_0[q_{i\alpha_j,\alpha_k}(\mathbf{Z}_i;\boldsymbol{\theta}_0)]
\end{aligned}$$

e

$$E_0[q_{i\beta_j,\beta_k}(\mathbf{Z}_i;\boldsymbol{\theta}_0)] = x_{i0}^2 E_0[q_{i\alpha_j,\alpha_k}(\mathbf{Z}_i;\boldsymbol{\theta}_0)] + \frac{4}{q\|\mathbf{b}_0\|^2} b_{jk}^0 E_0[W_f^2(\|\mathbf{e}_1\|^2)a(\|\mathbf{e}_1\|^2)\|\mathbf{e}_1\|^2],$$

uma vez que

$$E_0[q_{i\alpha_j}(\mathbf{Z}_i;\boldsymbol{\theta}_0)] = E_0[q_{i\beta_j}(\mathbf{Z}_i;\boldsymbol{\theta}_0)] = 0. \tag{3.79}$$

e assim a demonstração é completa usando a relação (3.13). □

Para provar os principais resultados desta seção precisamos das seguintes condições:

B3. Existem constantes positivas $a_f$ e $b_f$ tais que

$$|W_f(u)| \leq a_f \quad \text{e} \quad |W_{f'}(u)| \leq b_f, \quad u \geq 0;$$

B4. Para cada $\overline{\boldsymbol{\theta}} \in \Theta^0$ (interior de $\boldsymbol{\Theta}$) existem $\delta > 0$ e funções $d_i(\mathbf{Z}_i)$ e $d_{ikj}(\mathbf{Z}_i)$ tais que

$$|h_i(\mathbf{Z}_i; \boldsymbol{\theta})| \leq d_i(\mathbf{Z}_i) \quad \text{e} \quad |q_{i\theta_k\theta_j}(\mathbf{Z}_i; \boldsymbol{\theta})| \leq d_{ijk}(\mathbf{Z}_i),$$

para todo $\boldsymbol{\theta} \in B(\overline{\boldsymbol{\theta}}, \delta) = \{\boldsymbol{\theta}; \|\overline{\boldsymbol{\theta}} - \boldsymbol{\theta}\| < \delta\} \subset \boldsymbol{\Theta}^0$ e

$$\limsup \frac{1}{n} \sum_{i=1}^{n} E_0[d_i^2(\mathbf{Z}_i)] < \infty, \quad \limsup \frac{1}{n} \sum_{i=1}^{n} E_0[d_{ikj}^2(\mathbf{Z}_i)] < \infty,$$

$k, j = 1, \ldots, 2q$;

B5. Existe $\gamma > 0$ tal que

$$E_0[|W_f(\|\mathbf{e}_1\|^2)\|\mathbf{e}_1\| \, |u_1| \, |^{2+\gamma}] < \infty;$$

B6. A seqüência $x_1, x_2, \ldots$ é tal que existe uma constante positiva $M$ com

$$\sup |x_i| \leq M;$$

B7. Dado $\varepsilon > 0$, existe $\delta > 0$ tal que

$$\limsup \frac{1}{n} \left| \sum_{i=1}^{n} E_0[\sup\{q_{i\theta_k\theta_j}(\mathbf{Z}_i; \boldsymbol{\theta}); \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| < \delta\} - q_{i\theta_k\theta_j}(\mathbf{Z}_i; \boldsymbol{\theta}_0)] \right| < \varepsilon.$$

O mesmo é verdadeiro quando o sup é substituído pelo inf.

Mostraremos a seguir que o parâmetro $\boldsymbol{\theta}_0 = (\boldsymbol{\alpha}_0^\top, \boldsymbol{\beta}_0^\top)^\top$ tem a mesma propriedade que no modelo funcional univariado.

**Teorema 3.5** *Suponhamos que $B_f = 0$ e $C_f < 0$ e as condições B1 - B3 são satisfeitas. Então, $\boldsymbol{\theta}_0$ é um ponto de máximo local da função $\overline{\psi}(\boldsymbol{\theta}) = n^{-1}\sum_{i=1}^{n} E_0[h_i(\mathbf{Z}_i; \boldsymbol{\theta})]$, onde $h_i(\mathbf{Z}_i; \boldsymbol{\theta}) = \log f(\|\mathbf{T}_i\|^2)$.*

**Prova:** Para qualquer $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, segue de (3.69) que

$$q_{i\alpha_j}(\mathbf{Z}_i; \boldsymbol{\theta}) = -2W_f(\|\mathbf{T}_i\|^2)(\mathbf{d}_j^\top \mathbf{T}_i)$$

e

$$q_{i\beta_j}(\mathbf{Z}_i;\boldsymbol{\theta}) = q_{i\alpha_j}(\mathbf{Z}_i;\boldsymbol{\theta})x_{i0} - \frac{2}{\|\mathbf{b}\|}W_f(\|\mathbf{T}_i\|^2)(\mathbf{d}_j^\top\mathbf{T}_i)r_i.$$

Para $\boldsymbol{\theta} \in B(\boldsymbol{\theta}_0,\delta)$, $\delta > 0$, temos que

$$\|\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_{i0}\| \leq \|\mathbf{Z}_i - \mathbf{a}_0 - \mathbf{b}_0 x_{i0}\| + \delta(1+|x_{i0}|) = d_0(\mathbf{Z}_i),$$

o que implica que

$$\|\mathbf{T}_i\| = \|\mathbf{B}_2(\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_{i0})\| \leq \sqrt{q}\mathrm{d}_0(\mathbf{Z}_i) \quad \text{e} \quad |r_i| \leq \mathrm{d}_0(\mathbf{Z}_i).$$

Assim, da condição A3 juntamente com a desigualdade de Cauchy-Schwarz, segue que

$$|q_{i\alpha_j}(\mathbf{Z}_i;\boldsymbol{\theta})| \leq 2\sqrt{q}a_f\mathrm{d}_0(\mathbf{Z}_i), \quad j = 1,\dots,q,$$

e

$$|q_{i\beta_j}(\mathbf{Z}_i;\boldsymbol{\theta})| \leq 2\sqrt{q}a_f(\mathrm{d}_0(\mathbf{Z}_i)|x_{i0}| + d_0^2(\mathbf{Z}_i)), \quad j = 1,\dots,q.$$

Definindo

$$\psi_i(\boldsymbol{\theta}) = E_0[h_i(\mathbf{Z}_i;\boldsymbol{\theta})]$$

e usando as desigualdade acima e o Teorema da Convergência Dominada, temos que

$$\frac{\partial\psi_i(\boldsymbol{\theta})}{\partial\theta_j} = E_0[q_{i\theta_j}(\mathbf{Z}_i;\boldsymbol{\theta})], \quad j = 1,\dots,2q.$$

Assim, de (3.79), segue que $\boldsymbol{\theta}_0$ é um ponto crítico da função $\psi_i(\boldsymbol{\theta})$, $i = 1,\dots,n$, e conseqüentemente um ponto crítico de $\overline{\psi}(\boldsymbol{\theta})$, pois $\overline{\psi}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^n \psi_i(\boldsymbol{\theta})$. Por outro lado, da condição B3 e da desigualdade de Cauchy-Schwarz, temos que

$$|q_{i\alpha_k\alpha_j}(\mathbf{Z}_i;\boldsymbol{\theta})| \leq 4qc_f d_0^2(\mathbf{Z}_i) + 2qa_f,$$

$$|q_{i\alpha_k\beta_j}(\mathbf{Z}_i;\boldsymbol{\theta})| \leq 4qc_f(d_0^2(\mathbf{Z}_i)|x_{i0}| + d_0^3(\mathbf{Z}_i)) + 2qa_f(|x_{i0}| + 2d_0(\mathbf{Z}_i))$$

e

$$|q_{i\beta_k\beta_j}(\mathbf{Z}_i;\boldsymbol{\theta})| \leq 4qc_f(d_0(\mathbf{Z}_i)|x_{i0}| + d_0^2(\mathbf{Z}_i))^2 + 2qa_f(x_{i0}^2 + 4d_0(\mathbf{Z}_i)|x_{i0}| + 4d_0^2(\mathbf{Z}_i)),$$

$j, k = 1, \ldots, q$, onde $c_f = a_f(a_f + b_f)$. Assim, da condição B2 e do Teorema da Convergência Dominada, segue que

$$\frac{\partial^2}{\partial\theta_k \partial\theta_j}\psi_i(\boldsymbol{\theta}) = E_0[q_{i\theta_k\theta_j}(\mathbf{Z}_i; \boldsymbol{\theta})], \quad j, k = 1, \ldots, 2q,$$

e $\psi_i \in C^{(2)}$. Além disso, a matriz das segundas derivadas da função $\psi_i(\boldsymbol{\theta})$ em $\boldsymbol{\theta}_0$ é dada pela matriz $E_0[A_i(\mathbf{Z}_i; \boldsymbol{\theta}_0)]$ definida por

$$E_0[A_i(\mathbf{Z}_i; \boldsymbol{\theta}_0)] = 2C_f \begin{pmatrix} 1 & x_{i0} \\ x_{i0} & x_{i0}^2 + \frac{1}{||\mathbf{b}||^2}\frac{B_f}{C_f} \end{pmatrix} \otimes \mathbf{B}_{30},$$

com $C_f$, $B_f$ e $\mathbf{B}_{30}$ como no Lema 4.3. Conseqüentemente, a matriz das segundas derivadas de $\overline{\psi}(\boldsymbol{\theta})$ é dada pela matriz $E_0[A_n(\boldsymbol{\theta}_0)]$ definida no Lema 4.3, com $B_f = 0$, isto é,

$$E_0[A_n(\boldsymbol{\theta}_0)] = 2C_f \begin{pmatrix} 1 & \overline{x}_0 \\ \overline{x}_0 & S_x^0 \end{pmatrix} \otimes \mathbf{B}_{30}.$$

A seguir, mostraremos que os autovalores da matriz $E_0[A_n(\boldsymbol{\theta}_0)]$ são todos negativos e, conseqüentemente, $\boldsymbol{\theta}_0$ será um ponto de máximo local da função $\overline{\psi}(\boldsymbol{\theta})$. Com efeito, os autovalores da matriz

$$\mathbf{P}_0 = \begin{pmatrix} 1 & \overline{x}_0 \\ \overline{x}_0 & S_x^0 \end{pmatrix} \tag{3.80}$$

são dados por $\lambda_i = (1 + S_x^0) \pm \sqrt{(1 + S_x^0)^2 - 4S_{xx}^0}$, $i = 1, 2$, e são todos positivos (veja o caso univariado). Por outro lado,

$$\begin{aligned} |\mathbf{B}_{30} - \delta\mathbf{I}_q| &= \left|\mathbf{I}_q - \frac{\boldsymbol{\beta}_0\boldsymbol{\beta}_0^\top}{||\mathbf{b}_0||^2} - \delta\mathbf{I}_q\right| = \left|(1-\delta)\mathbf{I}_q - \frac{\boldsymbol{\beta}_0\boldsymbol{\beta}_0^\top}{||\mathbf{b}_0||^2}\right| \\ &= \left(\frac{||\boldsymbol{\beta}_0||^2}{||\mathbf{b}_0||^2}\right)^q |\gamma\mathbf{I}_q - \mathbf{B}_0|, \end{aligned}$$

onde

$$\gamma = \frac{||\mathbf{b}_0||^2}{||\boldsymbol{\beta}_0||^2}(1 - \delta) \quad \text{e} \quad \mathbf{B}_0 = \frac{\boldsymbol{\beta}_0\boldsymbol{\beta}_0^\top}{||\boldsymbol{\beta}_0||^2}. \tag{3.81}$$

Como $\mathbf{B}_0$ é uma matriz simétrica ~~independente~~ IDEMPOTENTE com posto 1, então os autovalores de $\mathbf{B}_0$ são $\gamma_1 = 1$, $\gamma_2 = \cdots = \gamma_q = 0$. Logo, de (3.81) segue que os autovalores de $\mathbf{B}_{30}$ são positivos e dados por $\delta_1 = \frac{1}{||\mathbf{b}_0||^2}$, $\delta_2 = \cdots = \delta_q = 1$ e, conseqüentemente, os autovalores de $\mathbf{P}_0 \otimes \mathbf{B}_{30}$ são dados por $\lambda_i \delta_j$, $i = 1, 2$, $j = 1, \ldots, q$, e com todos positivos (veja Fang e Zhang, 1990). Assim, os autovalores de $E_0[A_n(\boldsymbol{\theta}_0)] = 2C_f(\mathbf{P}_0 \otimes \mathbf{B}_{30})$, são negativos, desde que $C_f < 0$. Portanto, $\boldsymbol{\theta}_0$ é um ponto de máximo local da função $\overline{\psi}(\boldsymbol{\theta})$.

O resultado que apresentamos a seguir mostra a existência de um estimador $\widetilde{\boldsymbol{\theta}}_n$ que maximiza a relação (3.7), o qual é obtido resolvendo-se a equação

$$\sum_{i=1}^{n} \mathbf{Q}_i(\mathbf{Z}_i; \boldsymbol{\theta}) = 0, \tag{3.82}$$

onde $\mathbf{Q}_i(\mathbf{Z}_i; \boldsymbol{\theta}) = (q_{i\alpha_1}(\mathbf{Z}_i; \boldsymbol{\theta}), \ldots, q_{i\alpha_q}(\mathbf{Z}_i; \boldsymbol{\theta}), q_{i\beta_1}(\mathbf{Z}_i; \boldsymbol{\theta}), \ldots, \mathbf{Q}_{i\beta_q}(\mathbf{Z}_i; \boldsymbol{\theta}))^\top$, $i = 1, \ldots, n$, e é consistente e assintoticamente normal.

**Teorema 3.6** *Considere o modelo definido em (3.5) satisfazendo as condições B1 - B7 com $B_f = 0$ e $C_f < 0$. Então, o estimador de máxima verossimilhança $\widetilde{\boldsymbol{\theta}}_n = (\widetilde{\alpha}_n^\top, \widetilde{\boldsymbol{\beta}}_n^\top)^\top$ de $\boldsymbol{\theta}_0$, obtido resolvendo-se a equação (3.82), é tal que $\widetilde{\boldsymbol{\theta}}_n \xrightarrow{P} \boldsymbol{\theta}_0$. Além disso, $\widetilde{\boldsymbol{\theta}}_n$ é assintoticamente normal com média $\boldsymbol{\theta}_0$ e matriz de covariâncias dada por*

$$\boldsymbol{\Sigma}_n = \frac{E_0[W_f^2(\|\mathbf{e}_1\|^2)\|\mathbf{e}_1\|^2]}{qC_f^2(S_{xx}^0)^2} \begin{pmatrix} \Delta(S_{xx}^0)^2 + K\overline{x}_0^2 & -K\overline{x}_0 \\ -K\overline{x}_0 & K \end{pmatrix} \otimes \mathbf{B}_{40}, \tag{3.83}$$

*onde $\Delta = \|b_0\|^2$, $K = \Delta S_{xx}^0 + A_f$ e $\mathbf{B}_{40} = \frac{1}{\|\mathbf{b}_0\|^2}(\mathbf{I}_q + \boldsymbol{\beta}_0\boldsymbol{\beta}_0^\top)$.*

**Prova:** Sob o modelo funcional elíptico satisfazendo as condições B1 - B7, é fácil mostrar usando a desigualdade de Cauchy-Schwarz que as condições de regularidade consideradas em Mak (1982) são satisfeitas. Assim, pelos resultados na Seção 3.2, onde é apresentado um breve resumo dos principais resultados de Mak (1982), segue que o estimador de máxima verossimilhança $\widetilde{\boldsymbol{\theta}}_n$, obtido resolvendo-se a equação (3.82), é consistente e assintoticamente normal com média $\boldsymbol{\theta}_0$ e onde a matriz de covariâncias segue de (3.15) e é dada por

$$\boldsymbol{\Sigma}_n = (E_0[A_n(\boldsymbol{\theta}_0)])^{-1} V_n(\boldsymbol{\theta}_0)(E_0[A_n(\boldsymbol{\theta}_0)])^{-1},$$

onde as matrizes $E_0[A_n(\boldsymbol{\theta}_0)]$ e $V_n(\boldsymbol{\theta}_0)$ são como no Lema 4.1 com $B_f = 0$, ou seja,

$$(E_0[A_n(\boldsymbol{\theta}_0)])^{-1} = \frac{1}{2C_f S_{xx}^0}\begin{pmatrix} S_x^0 & -\overline{x}_0 \\ -\overline{x}_0 & 1 \end{pmatrix} \otimes \mathbf{B}_{30}^{-1}, \tag{3.84}$$

$$V_n(\boldsymbol{\theta}_0) = \frac{4}{q} E_0[W_f^2(\|\mathbf{e}_1\|^2)\|\mathbf{e}_1\|^2] \begin{pmatrix} S_x^0 & \overline{x}_0 \\ \overline{x}_0 & \frac{1}{\|b_0\|^2} A_f \end{pmatrix} \otimes \mathbf{B}_{30}, \tag{3.85}$$

onde $\overline{x}_0$, $S_x^0$, $A_f$, $C_f$ e $\mathbf{B}_{30}$ são como no Lema 4.1. □

## 3.5 O modelo funcional $t$-Student multi-univariado

Nesta seção estudaremos as propriedades assintóticas do estimador de máxima verossimilhança do parâmetro estrutural do modelo estatístico definido em (3.5), quando a seqüência de vetores de erros $\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n$ tem distribuição comum $t$-Student, isto é, quando

$$\boldsymbol{\epsilon}_1 \sim t_p(\mathbf{0}, \mathbf{I}_p; \nu), \tag{3.86}$$

cuja função de densidade é dada por $f(\mathbf{x}^\top \mathbf{x})$, $\mathbf{x} \in \mathbb{R}^p$, com $p = q + 1$ e

$$f(u) = k(p, \nu)\nu^{\nu/2}(\nu + u)^{-(\nu+p)/2}, \; u \geq 0, \tag{3.87}$$

onde

$$k(p, \nu) = \frac{\Gamma(\frac{1}{2}(\nu + p))}{\pi^{\frac{1}{2}p}\Gamma(\frac{1}{2}\nu)}. \tag{3.88}$$

De (3.87) é fácil ver que

$$W_f(u) = -\frac{\nu + p}{2}(\nu + u)^{-1} \quad \text{e} \quad \Delta_f(u) = \frac{\nu + p}{2}(\nu + u)^{-2}, \; u \geq 0. \tag{3.89}$$

Seja $\mathbf{X} \sim t_p(\mathbf{0}, \mathbf{I}_p; \lambda, \nu)$ a distribuição $t$-Student generalizada (veja apêndice), com densidade $f_p(\mathbf{x}^\top \mathbf{x})$, $\mathbf{x} \in \mathbb{R}^p$, onde

$$f_p(u) = k(\nu, p)\lambda^{\nu/2}(\lambda + u)^{-(\nu+p)/2}, \; u \geq 0. \tag{3.90}$$

O índice $p$ e $f_p$ é para indicar que $f_p(\mathbf{x}^\top \mathbf{x})$ é a densidade do vetor aleatório $p$-dimensional $\mathbf{X} \sim t_p(\mathbf{0}, \mathbf{I}_p; \lambda, \nu)$.

Definimos as funções $W_p$ e $\Delta_p(u)$ por

$$W_p(u) = \frac{f_p'(u)}{f_p(u)} \quad \text{e} \quad \Delta_p(u) = W_p'(u), \; u \geq 0, \tag{3.91}$$

onde

$$f_p'(u) = \frac{\mathrm{d}}{\mathrm{d}u} f_p(u) \quad \text{e} \quad W_p'(u) = \frac{\mathrm{d}}{\mathrm{d}u} W_f(u).$$

Seja $\mathbf{X} = (\mathbf{X}_1^\top, \mathbf{X}_2^\top)^\top \sim t_p(\mathbf{0}, \mathbf{I}_p; \lambda, \nu)$ com densidade como em (3.90) e $\mathbf{X}_1$ o vetor $q$-dimensional, $1 \leq q < p$ com densidade marginal $f_q(\mathbf{y}^\top \mathbf{y})$, $\mathbf{y} \in \mathbb{R}^q$. Para esta última densidade, seja $W_q(u) = f_q'(u)/f_q(u)$, $u \geq 0$.

**Lema 3.8** *Considere a distribuição t-Student generalizada definida em (3.90). Então,*

*i)* $W_p(u) = -\dfrac{\nu + p}{2}(\lambda + u)^{-1}$, $u \geq 0$;

*ii)* $W_p(u) = \dfrac{\nu + p}{\nu + q} W_q(u)$, $u \geq 0$, $1 \leq q < p$;

*iii)* $W_p'(u) = 2\dfrac{\nu + p}{(\nu + q)^2} W_q^2(u)$.

**Lema 3.9** *Seja* $\mathbf{X} = (\mathbf{X}_1^\top, \mathbf{X}_2^\top)^\top \sim t_p(\mathbf{0}, \mathbf{I}_p; \lambda, \nu)$, *com* $\mathbf{X}_1$ *vetor* $q$*-dimensional. Então,*

*i)* $E[\|\mathbf{X}\|^{2k} W_p(\|\mathbf{X}\|^2)^m] = \left(\dfrac{\nu + p}{2}\right)^m \dfrac{B\left[\frac{1}{2}(p + 2k), \frac{1}{2}(\nu + 2m - 2k)\right]}{\lambda^{(2m-2k)/2} B\left[\frac{1}{2}p, \frac{1}{2}\nu\right]}$,

*ii)* $E[\|\mathbf{X}_1\|^{2k} (W_p(\|\mathbf{X}_1\|^2)^m] = \left(\dfrac{\nu + p}{2}\right)^m \dfrac{B\left[\frac{1}{2}(q + 2k), \frac{1}{2}(\nu + 2m - 2k)\right]}{\lambda^{(2m-2k)/2} B\left[\frac{1}{2}q, \frac{1}{2}\nu\right]}$.

**Prova:** A demonstração de i) segue facilmente do Lema A.1, dado no Apêndice A (veja também Arellano-Valle e Bolfarine, 1995). A prova de ii) segue de (ii) do Lema 3.8 e da parte i). □

O seguinte resultado será de muita utilidade para determinar as matrizes $V_n(\boldsymbol{\theta}_0)$ e $E_0[A_n(\boldsymbol{\theta}_0)]$ definidas em (3.12) e (3.13), respectivamente.

**Lema 3.10** *Seja* $\mathbf{X} = (\mathbf{X}_1^\top, X_2)^\top \sim t_p(\mathbf{0}, \mathbf{I}_p; \lambda, \nu)$, $p = q + 1$, *com densidade dada por (3.90) e* $\mathbf{X}_1$ *de dimensão* $q$. *Então,* *1)* $E[W_p(\|\mathbf{X}_1\|^2)] = -\frac{1}{2}\frac{\nu(\nu+p)}{\lambda(\nu+q)}$;

*2)* $E[W_p(\|\mathbf{X}_1\|^2)\|\mathbf{X}_1\|^2] = -\frac{q}{2}\frac{\nu+p}{\nu+q}$;

*3)* $E[W_p^2(\|\mathbf{X}_1\|^2)\|\mathbf{X}_1\|^2] = \frac{q}{4}\frac{\nu(\nu+p)}{\lambda(\nu+q)(\nu+q+2)}$;

*4)* $E[\Delta_p(\|\mathbf{X}_1\|^2)\|\mathbf{X}_1\|^2] = \frac{q}{2}\frac{\nu(\nu+p)}{\lambda(\nu+q)(\nu+q+2)}$;

*5)* $E[W_p(\|\mathbf{X}_1\|^2)X_2^2] = -\frac{1}{2}\frac{\nu+p}{\nu+q-2}$;

*6)* $E[W_p^2(\|\mathbf{X}_1\|^2)\|\mathbf{X}_1\|^2 X_2^2)] = \frac{q}{4}\frac{(\nu+p)^2}{(\nu+q)(\nu+q-2)}$;

*7)* $E[\Delta_p(\|\mathbf{X}_1\|^2)\|\mathbf{X}_1\|^2 X_2^2)] = \frac{q}{2}\frac{\nu+p}{(\nu+q)(\nu+q-2)}$,

*onde* $\Delta_p(u) = \frac{\mathrm{d}}{\mathrm{d}u}W_p(u)$.

**Prova:** Do Lema 3.9 segue que

$$E[W_p^2(\|\mathbf{X}_1\|^2)\|\mathbf{X}_1\|^2)] = -\frac{q}{2}, \quad E[W_q^2(\|\mathbf{X}_1\|^2)\|\mathbf{X}_1\|^2] = \frac{q}{4}\frac{\nu(\nu+q)}{\lambda(\nu+q+2)}$$

$$\text{e} \qquad E[W_p(\|\mathbf{X}_1\|^2)] = -\frac{1}{2}\frac{\nu}{\lambda} \tag{3.92}$$

e de iii) da seção A.3 do Apêndice A, segue que

$$E[X_2^2|\mathbf{X}_1] = \frac{\lambda+\|\mathbf{X}_1\|^2}{\nu+q-2} = -\frac{1}{2}\frac{\nu+q}{\nu+q-2}W_q^{-1}(\|\mathbf{X}_1\|^2). \tag{3.93}$$

A seguir provaremos as afirmações em 6) e 7). As demonstrações de 1), 2), 3), 4) e 5) são análogas. Com efeito, usando o Lema 3.8 juntamente com (3.92) e (3.93), temos que

$$\begin{aligned}
E[W_p^2(\|\mathbf{X}_1\|^2)\|\mathbf{X}_1\|^2 X_2^2] &= \left(\frac{\nu+p}{\nu+q}\right)^2 E[W_q^2(\|\mathbf{X}_1\|^2)\|\mathbf{X}_1\|^2 X_2^2] \\
&= \left(\frac{\nu+p}{\nu+q}\right)^2 E[W_q^2(\|\mathbf{X}_1\|^2)\|\mathbf{X}_1\|^2 E[X_2^2|\mathbf{X}_1]] \\
&= \left(\frac{\nu+p}{\nu+q}\right)^2 E\left[W_q^2(\|\mathbf{X}_1\|^2)\|\mathbf{X}_1\|^2\left(-\frac{1}{2}\frac{\nu+q}{\nu+q-2}W_q^{-1}(\|\mathbf{X}_1\|^2)\right)\right] \\
&= -\frac{1}{2}\frac{(\nu+p)^2}{(\nu+q)(\nu+q-2)}E[W_q(\|\mathbf{X}_1\|^2)\|\mathbf{X}_1\|^2] \\
&= \frac{q}{4}\frac{(\nu+p)^2}{(\nu+q)(\nu+q-2)}.
\end{aligned}$$

A prova de 7) é similar. De fato, usando o Lema 3.8, (3.92) e (3.93), temos que

$$
\begin{aligned}
E[\Delta_f(\|\mathbf{X}_1\|^2)\|\mathbf{X}_1\|^2 X_2^2] &= 2\frac{\nu+p}{\nu+q} E[W_q^2(\|\mathbf{X}_1\|^2)\|\mathbf{X}_1\|^2 X_2^2] \\
&= -\frac{\nu+p}{(\nu+q)(\nu+q-2)} E[W_q(\|\mathbf{X}_1\|^2)\|\mathbf{X}_1\|^2] \\
&= \frac{q}{2}\frac{\nu+p}{(\nu+q)(\nu+q-2)}.
\end{aligned}
$$

□

Do Lema 3.10, com $\lambda = \nu$ e $W_p = W_f$, onde $f$ é como em (3.87), segue que

$$
\begin{aligned}
& E_0[W_f(\|\mathbf{e}_1\|)] = -\frac{1}{2}\frac{\nu+p}{\nu+q}; \quad E_0[W_f(\|\mathbf{e}_1\|^2)\|\mathbf{e}_1\|^2] = -\frac{q}{2}\frac{\nu+p}{\nu+q}; \\
& E_0[W_f^2(\|\mathbf{e}_1\|^2)\|\mathbf{e}_1\|^2] = \frac{q}{4}\frac{(\nu+p)^2}{(\nu+q)(\nu+q+2)}; \\
& E_0[\Delta_f(\|\mathbf{e}_1\|^2)\|\mathbf{e}_1\|^2] = \frac{q}{2}\frac{\nu+p}{(\nu+q)(\nu+q+2)}; \\
& E_0[W_f^2(\|\mathbf{e}_1\|^2)a(\|\mathbf{e}_1\|^2)] = -\frac{1}{2}\frac{\nu+p}{\nu+q-2}; \\
& E_0[W_f(\|\mathbf{e}_1\|^2)a(\|\mathbf{e}_1\|^2)\|\mathbf{e}_1\|^2] = \frac{q}{4}\frac{(\nu+p)^2}{(\nu+q)(\nu+q-2)} \\
\text{e} \quad & E_0[\Delta_f(\|\mathbf{e}_1\|^2)a(\|\mathbf{e}_1\|^2)\|\mathbf{e}_1\|^2] = \frac{q}{2}\frac{\nu+p}{(\nu+q)(\nu+q-2)},
\end{aligned}
$$

onde $a(\|\mathbf{e}_1\|^2) = E_0[u_1^2|\mathbf{e}_1] = \dfrac{\nu+\|\mathbf{e}_1\|^2}{\nu+q-2}$ e $(\mathbf{e}_1^\top, u_1) \sim t_p(\mathbf{0}, \mathbf{I}_p; \nu)$, $\nu > 1$, $p = q+1$ e $\mathbf{e}_1$ sendo um vetor aleatório $q$-dimensional. Logo, após algumas manipulações algébricas, segue que os coeficientes $A_f$, $B_f$ e $C_f$, definidos no Lema 3.7, são dados por

$$
A_f = \frac{\nu+q+2}{\nu+q-2}, \quad B_f = 0 \quad \text{e} \quad C_f = -\frac{1}{2}\frac{\nu+p}{\nu+q+2}. \tag{3.94}
$$

Conseqüentemente, as matrizes $V_n(\boldsymbol{\theta}_0)$ e $E_0[A_n(\boldsymbol{\theta}_0)]$ são dadas por

$$
V_n(\boldsymbol{\theta}_0) = \frac{(\nu+p)^2}{(\nu+q)(\nu+q+2)} \begin{pmatrix} 1 & \overline{x}_0 \\ \overline{x}_0 & S_x^0 + \frac{1}{\|b_0\|^2} A_f \end{pmatrix} \otimes \mathbf{B}_{30} \tag{3.95}
$$

$$
E_0[A_n(\boldsymbol{\theta}_0)] = -\frac{\nu+p}{\nu+q+2} \begin{pmatrix} 1 & \overline{x}_0 \\ \overline{x}_0 & S_x^0 \end{pmatrix} \otimes \mathbf{B}_{30}, \tag{3.96}
$$

onde $\overline{x}_0$, $S_x^0$ e $\mathbf{B}_{30}$ são como no Lema 3.7.

Dos Lemas 3.8, 3.9 e 3.10 é fácil ver que $B_f = 0$ e $C_f < 0$, quando $\boldsymbol{\epsilon}_i \overset{\text{iid}}{\sim} t_p(\mathbf{0}, \mathbf{I}_p; \lambda, \nu)$, isto é, quando os vetores de erros $\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n$ têm distribuição comum $t$-Student generalizada. Portanto, também valem os teoremas desta seção, dados no contexto da distribuição $t$-Student.

De (3.82) temos, após algumas manipulações algébricas, que o estimador de máxima verossimilhança de $\boldsymbol{\theta}_0 = (\boldsymbol{\alpha}_0^\top, \boldsymbol{\beta}_0^\top)^\top$ é obtido resolvendo-se iterativamente a equação em (3.82), que no caso do modelo funcional $t$-Student podem ser escrita como

$$\sum_{i=1}^{n} \frac{\mathbf{d}_j^\top \mathbf{T}_i}{\nu + \|\mathbf{T}_i\|^2} \begin{pmatrix} 1 \\ x_{i0} + \frac{r_i}{\|\mathbf{b}\|} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{3.97}$$

$j = 1, \ldots, q$, já que

$$q_{i\alpha j}(\mathbf{Z}_i; \boldsymbol{\theta}) = (\nu + p)(\nu + \|\mathbf{T}_i\|^2)^{-1}(\mathbf{d}_j^\top \mathbf{T}_i)$$

e

$$q_{i\beta j}(\mathbf{Z}_i; \boldsymbol{\theta}) = (\nu + p)(\nu + \|\mathbf{T}_i\|^2)^{-1}\mathbf{d}_j^\top \mathbf{T}_i(x_{i0} - \frac{r_i}{\|\mathbf{b}\|}),$$

$i = 1, \ldots, n$, donde $\mathbf{T}_i$ e $r_i$ são como em (3.65). Contudo, Kimura (1992) sugere usar um algoritmo do tipo EM para obter o estimador de máxima verossimilhança do parâmetro estrutural do modelo funcional definido em (3.5). A distribuição $t$-Student é estudada em Arellano-Valle e Bolfarine (1995).

**Teorema 3.7** *Considere o modelo definido em (3.5) e (3.87) satisfazendo a condição B6. Então, para $\nu > 4$, o estimador de máxima verossimilhança $\widetilde{\boldsymbol{\theta}}_n = (\widetilde{\boldsymbol{\alpha}}_n^\top, \widetilde{\boldsymbol{\beta}}_n^\top)^\top$ de $\boldsymbol{\theta}_0 = (\boldsymbol{\alpha}_0^\top, \boldsymbol{\beta}_0^\top)^\top$, que é solução da equação (3.97), é tal que $\widetilde{\boldsymbol{\theta}}_0 \xrightarrow{\text{P}} \boldsymbol{\theta}_0$. Além disso, é assintoticamente normal com média $\boldsymbol{\theta}_0$ e matriz de covariâncias dada por*

$$\boldsymbol{\Sigma}_n = \frac{\nu + q + 2}{\nu + q} \frac{1}{(S_{xx}^0)^2} \begin{pmatrix} \Delta(S_{xx}^0)^2 + K\overline{x}_0^2 & -K\overline{x}_0 \\ -K\overline{x}_0 & K \end{pmatrix} \otimes \mathbf{B}_{40},$$

*onde $\Delta = \|b_0\|^2$, $K = \Delta S_{xx}^0 + A_f$, com $A_f$, $S_{xx}^0$ e $\mathbf{B}_{40}$ definidos como em (3.94) e no Teorema 3.6, respectivamente.*

**Prova:** De forma similar à do modelo fucional univariado, prova-se que as condições B1 - B7 são satisfeitas. Assim, o resultado segue do Teorema 4.2, juntamente com (3.95) e (3.96).

# Capítulo 4

# O teste de razão de verossimilhança e correções de Bartlett no modelo estrutural com mais de uma população

## 4.1 Introdução

Neste capítulo o objetivo é fazer inferência sobre os coeficientes angulares no modelo estatístico que consiste em $k$ modelos estruturais com erros nas variáveis, isto é, no modelo que é definido por

$$\begin{cases} Y_{ij} = \alpha_j + \beta_j x_{ij} + e_{ij} \\ X_{ij} = x_{ij} + u_{ij} \end{cases} \qquad i = 1, \ldots, n_j, \quad j = 1, \ldots, k. \tag{4.1}$$

Com tal objetivo, vamos supor, em princípio, que os erros aleatórios $e_{ij}$ e $u_{ij}$, e a variável aleatória $x_{ij}$ satisfazem as seguintes condições:

i) $E[e_{ij}] = 0$, $\mathrm{Var}[e_{ij}] = \sigma^2_{e_j}$, $\mathrm{Cov}[e_{i'j'}, e_{ij}] = 0$, $(i', j') \neq (i, j)$;

ii) $E[u_{ij}] = 0$, $\mathrm{Var}[u_{ij}] = \sigma^2_{u_j}$, $\mathrm{Cov}[u_{i'j'}, u_{ij}] = 0$, $(i', j') \neq (i, j)$;

iii) $E[x_{ij}] = \mu_{x_j}$, $\mathrm{Var}[x_{ij}] = \sigma^2_{xj}$, $\mathrm{Cov}[x_{i'j'}, x_{ij}] = 0$, $(i', j') \neq (i, j)$;

iv) $\mathrm{Cov}[e_{i'j'}, u_{ij}] = \mathrm{Cov}[x_{i'j'}, e_{ij}] = \mathrm{Cov}[x_{i'j'}, u_{ij}] = 0$, $(i', j') \neq (i, j)$.

Com tais hipóteses o modelo definido em (4.1) representa uma extensão do modelo estrutural definido na seção 2.5 do Capítulo 2 e pode ser escrito por

$$\mathbf{Z}_{ij} = \mathbf{a}_j + \mathbf{B}_j \mathbf{r}_{ij}, \tag{4.2}$$

onde

$$\mathbf{Z}_{ij} = \begin{pmatrix} Y_{ij} \\ X_{ij} \end{pmatrix}, \quad \mathbf{r}_{ij} = \begin{pmatrix} x_{ij} \\ e_{ij} \\ u_{ij} \end{pmatrix}, \quad \mathbf{a}_j = \begin{pmatrix} \alpha_j \\ 0 \end{pmatrix}, \quad \mathbf{B}_j = [\mathbf{b}_j, I_2] \quad \text{e} \quad \mathbf{b}_j = \begin{pmatrix} \beta_j \\ 1 \end{pmatrix}.$$

$i = 1, \ldots, n_j$, $j = 1, \ldots, k$. Assim, de (4.2), segue que a distribuição de $\mathbf{Z}_{ij}$ é determinada pela distribuição de $\mathbf{r}_{ij}$.

Para efeito de inferência, encontramos certos inconvenientes que deficultam a estimação consistente dos parâmetros de interesse do modelo. Neste caso, o modelo pode ser não identificável, no sentido de que valores diferentes do vetor de parâmetros podem dar origem à mesma distribuição das observações. Isto acontece, por exemplo, com o modelo normal (veja Wong, 1991, e Nascimento, 1994), a menos de uma suposição adicional (além das dadas em i) a iv)) que permita reduzir o número de parâmetros envolvidos no modelo. Sob normalidade, são consideradas suposições sobre as variâncias de $x_{ij}$, $e_{ij}$ e $u_{ij}$ (Wong, 1991, e Nascimento, 1994).

O objetivo deste capítulo é estender alguns resultados obtidos no modelo normal para modelos elípticos. Discutiremos a inferência dos parâmetros $\beta_1, \ldots, \beta_k$, sob o modelo normal usando o método da razão de máxima verossimilhança. Correções de Bartlett para diferentes hipóteses são obtidas, avaliando diretamente o valor esperado da estatística da razão de máxima verossimilhança sob a hipótese nula, que é um método alternativo à metodologia de Lawley (1956). Extensões são consideradas para a classe das distribuições elípticas.

## 4.2 Inferência no modelo estrutural normal supondo razões de variâncias conhecidas

Nesta seção estudaremos o modelo estrutural normal, isto é, vamos supor que em (4.2),

$$\mathbf{r}_{ij} \overset{\text{ind}}{\sim} N_3(\boldsymbol{\eta}_j, \boldsymbol{\Omega}_j), \quad i = 1, \ldots, n_j, \tag{4.3}$$

onde

$$\boldsymbol{\eta}_j = \begin{pmatrix} \mu_{x_j} \\ 0 \\ 0 \end{pmatrix} \quad \text{e} \quad \boldsymbol{\Omega}_j = \begin{pmatrix} \sigma^2_{x_j} & 0 & 0 \\ 0 & \sigma^2_{e_j} & 0 \\ 0 & 0 & \sigma^2_{u_j} \end{pmatrix}.$$

Logo, a partir de (4.2), é claro que

$$\mathbf{Z}_{ij} \overset{\text{ind}}{\sim} N_2(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j), \quad i = 1, \ldots, n_j, \tag{4.4}$$

onde $\boldsymbol{\mu}_j = \mu(\boldsymbol{\theta}_j) = E[\mathbf{Z}_{ij}]$ e $\boldsymbol{\Sigma}_j = \Sigma(\boldsymbol{\theta}_j) = \text{Var}[\mathbf{Z}_{ij}]$ são dados por

$$\boldsymbol{\mu}_j = \mathbf{a}_j + \mathbf{B}_j\boldsymbol{\eta}_j = \begin{pmatrix} \alpha_j + \beta_j\mu_{x_j} \\ \mu_{x_j} \end{pmatrix} \quad \text{e} \quad \boldsymbol{\Sigma}_j = \mathbf{B}_j\boldsymbol{\Omega}_j\mathbf{B}_j^\top = \begin{pmatrix} \beta_j^2\sigma^2_{x_j} + \sigma^2_{e_j} & \beta_j\sigma^2_{x_j} \\ \beta_j\sigma^2_{x_j} & \sigma^2_{x_j} + \sigma^2_{u_j} \end{pmatrix}, \tag{4.5}$$

com

$$\boldsymbol{\theta}_j = (\alpha_j, \mu_{x_j}, \sigma^2_{x_j}, \sigma^2_{e_j}, \sigma^2_{u_j}, \beta_j)^\top. \tag{4.6}$$

Assim, para uma amostra de tamanho $n_j$, da $j$-ésima população, $j = 1, \ldots, k$, a função de verossimilhança é dada por

$$L_j(\mathbf{Z}_{1j}, \ldots, \mathbf{Z}_{n_j j}|\boldsymbol{\theta}_j) = |\boldsymbol{\Sigma}_j|^{-n_j/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n_j}(\mathbf{Z}_{ij} - \boldsymbol{\mu}_j)^\top\boldsymbol{\Sigma}_j^{-1}(\mathbf{Z}_{ij} - \boldsymbol{\mu}_j)\right),$$

com logaritmo dado por

$$L_j(\boldsymbol{\theta}_j) = -\frac{n_j}{2}\log|\boldsymbol{\Sigma}_j| - \frac{1}{2}\sum_{i=1}^{n_j}(\mathbf{Z}_{ij} - \boldsymbol{\mu}_j)^\top\boldsymbol{\Sigma}_j^{-1}(\mathbf{Z}_{ij} - \boldsymbol{\mu}_j). \tag{4.7}$$

Portanto, para uma amostra de tamanho $n = n_1 + \cdots + n_k$, do modelo em (4.2), o logaritmo da função de verossimilhança é dado por

$$\begin{aligned} L(\boldsymbol{\theta}) &= \sum_{j=1}^{k} L_j(\boldsymbol{\theta}_j) \\ &= \sum_{j=1}^{k}\left(-\frac{n_j}{2}\right)\log|\boldsymbol{\Sigma}_j| - \frac{1}{2}\sum_{j=1}^{k}\sum_{i=1}^{n_j}(\mathbf{Z}_{ij} - \boldsymbol{\mu}_j)^\top\boldsymbol{\Sigma}_j^{-1}(\mathbf{Z}_{ij} - \boldsymbol{\mu}_j), \end{aligned} \tag{4.8}$$

onde $\boldsymbol{\mu}_j$ e $\boldsymbol{\Sigma}_j$ são como em (4.5) e $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \ldots, \boldsymbol{\theta}_k^\top)^\top$, com $\boldsymbol{\theta}_j$ como em (4.6).

Similarmente ao caso de uma população, definimos

$$\overline{\mathbf{Z}}_j = \frac{1}{n_j}\sum_{i=1}^{n_j}\mathbf{Z}_{ij} \quad \text{e} \quad \mathbf{S}_{n_j} = \frac{1}{n_j}\sum_{i=1}^{n_j}(\mathbf{Z}_{ij} - \overline{\mathbf{Z}}_j)(\mathbf{Z}_{ij} - \overline{\mathbf{Z}}_j)^\top, \tag{4.9}$$

como a média e a matriz de covariâncias amostrais, respectivamente, da $j$-ésima população. Neste caso, note que

$$\mathbf{S}_{n_j} = \begin{pmatrix} S_{YY}^{(j)} & S_{XY}^{(j)} \\ S_{XY}^{(j)} & S_{XX}^{(j)} \end{pmatrix}, \tag{4.10}$$

onde

$$S_{YY}^{(j)} = \frac{1}{n_j}\sum_{i=1}^{n_j}(Y_{ij} - \overline{Y}_j)^2, \qquad S_{XY}^{(j)} = \frac{1}{n_j}\sum_{i=1}^{n_j}(X_{ij} - \overline{X}_j)(Y_{ij} - \overline{Y}_j)$$

e

$$S_{XX}^{(j)} = \frac{1}{n_j}\sum_{i=1}^{n_j}(X_{ij} - \overline{X}_j)^2,$$

$j = 1, \ldots, k$. Assim, a média e a matriz de covariâncias amostrais da amostra completa (4.2) (modelo consistindo de $k$ modelos estruturais) são dadas por

$$\overline{\mathbf{Z}} = \frac{1}{n}\sum_{j=1}^{k} n_j\overline{\mathbf{Z}}_j \quad \text{e} \quad \mathbf{S} = \frac{1}{n}\sum_{j=1}^{k} n_j\mathbf{S}_{n_j}, \tag{4.11}$$

respectivamente, onde $n = n_1 + \cdots + n_k$. Assim, a expressão dada em (4.8) pode ser escrita como

$$L(\boldsymbol{\theta}) = \sum_{j=1}^{k}\left(-\frac{n_j}{2}\right)\log|\boldsymbol{\Sigma}_j| - \frac{1}{2}\sum_{j=1}^{k} n_j\left[\operatorname{tr}(\boldsymbol{\Sigma}_j^{-1}\mathbf{S}_{n_j}) + (\overline{\mathbf{Z}}_j - \boldsymbol{\mu}_j)^\top\boldsymbol{\Sigma}_j^{-1}(\overline{\mathbf{Z}}_j - \boldsymbol{\mu}_j)\right]. \tag{4.12}$$

Agora, na população $j$, $j = 1, \ldots, k$, o EMV de $\boldsymbol{\mu}_j$ e $\boldsymbol{\Sigma}_j$ podem ser obtidos como solução das equações de momentos

$$\widehat{\boldsymbol{\mu}}_j = \overline{\mathbf{Z}}_j \quad \text{e} \quad \widehat{\boldsymbol{\Sigma}}_j = \mathbf{S}_{n_j} \tag{4.13}$$

ou, equivalentemente,

$$\begin{pmatrix} \widehat{\alpha}_j + \widehat{\beta}_j\widehat{\mu}_{x_j} \\ \widehat{\mu}_{x_j} \end{pmatrix} = \begin{pmatrix} \overline{Y}_j \\ \overline{X}_j \end{pmatrix} \quad \text{e} \quad \begin{pmatrix} \widehat{\beta}_j\widehat{\sigma}_{x_j}^2 + \widehat{\sigma}_{e_j}^2 \\ \widehat{\beta}_j\widehat{\sigma}_{x_j}^2 \\ \widehat{\sigma}_{x_j}^2 + \widehat{\sigma}_{u_j}^2 \end{pmatrix} = \begin{pmatrix} S_{YY}^{(j)} \\ S_{XY}^{(j)} \\ S_{XX}^{(j)} \end{pmatrix},$$

$j = 1, \ldots, k$. Da relação acima, segue que não é possível obter o EMV de $\boldsymbol{\theta}_j$, pois somente $\mu_{x_j}$ está identificado. Para maiores detalhes, veja por exemplo, Kendall e Stuart (1979). Para tornar o modelo correspondente à $j$-ésima população identificável, a maioria dos autores concordam em colocar condições sobre as variâncias dos erros ($\sigma^2_{e_j}$ e $\sigma^2_{u_j}$) e não sobre os parâmetros de maior interesse, que neste caso são os parâmetros de regressão $\alpha_j$ e $\beta_j$.

Neste trabalho, vamos considerar as seguintes restrições:

$$\lambda_{e_j} = \sigma^2_{e_j}/\sigma^2_{u_j} \text{ ou } \lambda_{x_j} = \sigma^2_{x_j}/\sigma^2_{u_j} \tag{4.14}$$

conhecido (veja, por exemplo, Kendall e Stuart, 1979). Assim, para cada uma destas suposições, o vetor $\boldsymbol{\theta}_j$ definido em (4.6) torna-se identificável, já que se reduz a

$$\boldsymbol{\theta}_j = \begin{cases} (\alpha_j, \mu_{x_j}, \sigma^2_{e_j}, \sigma^2_{u_j}, \beta_j)^\top & \text{, se } \lambda_{x_j} \text{ é conhecido} \\ (\alpha_j, \mu_{x_j}, \sigma^2_{x_j}, \sigma^2_{u_j}, \beta_j)^\top & \text{, se } \lambda_{e_j} \text{ é conhecido,} \end{cases} \tag{4.15}$$

$j = 1, \ldots, k$. Como $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \ldots, \boldsymbol{\theta}_k^\top)^\top$, $\boldsymbol{\mu}_j = \boldsymbol{\mu}(\boldsymbol{\theta}_j)$ e $\boldsymbol{\Sigma}_j = \boldsymbol{\Sigma}(\boldsymbol{\theta}_j)$, $j = 1, \ldots, k$, então para ambas as condições de identificabilidade $\lambda_{x_j}$ ou $\lambda_{e_j}$, $j = 1, \ldots, k$, conhecidos, temos que a relação

$$\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \ldots, \boldsymbol{\theta}_k^\top)^\top \to (\boldsymbol{\mu}(\boldsymbol{\theta}_1), \boldsymbol{\Sigma}(\boldsymbol{\theta}_1), \ldots, \boldsymbol{\mu}(\boldsymbol{\theta}_k), \boldsymbol{\Sigma}(\boldsymbol{\theta}_k))$$

é um a um, e desta maneira tornamos $\boldsymbol{\theta}$ identificável. A condição $\lambda_{e_j}$ conhecido, $j = 1, \ldots, k$, é considerada por Wong (1991) para testar hipóteses sobre os coeficientes angulares $\beta_1, \ldots, \beta_k$. Um estudo similar, mas com a condição de identificabilidade $\lambda_{x_j}$ conhecido, foi considerada por Nascimento (1994). Esta última condição de identificabilidade é equivalente a estabelecer que

$$k_{x_j} = \frac{\lambda_{x_j}}{1 + \lambda_{x_j}}, \quad j = 1, \ldots, k,$$

são conhecidos (veja Capítulo 2).

Note que sob as condições de identificabilidade consideradas, segue que a matriz $\boldsymbol{\Sigma}_j = \text{Cov}[\mathbf{Z}_{ij}]$ pode ser escrita como

$$\boldsymbol{\Sigma}_j = \begin{cases} \begin{pmatrix} \lambda_{x_j}\beta_j^2\sigma_{u_j}^2 + \sigma_{e_j}^2 & \lambda_{x_j}\beta_j\sigma_{u_j}^2 \\ \lambda_{x_j}\beta_j\sigma_{u_j}^2 & (\lambda_{x_j}+1)\sigma_{u_j}^2 \end{pmatrix} & \text{, se } \lambda_{x_j} \text{ é conhecido} \\ \begin{pmatrix} \beta_j^2\sigma_{x_j}^2 + \lambda_{e_j}\sigma_{u_j}^2 & \beta_j\sigma_{x_j}^2 \\ \beta_j\sigma_{x_j}^2 & \sigma_{x_j}^2 + \sigma_{u_j}^2 \end{pmatrix} & \text{, se } \lambda_{e_j} \text{ é conhecido,} \end{cases} \tag{4.16}$$

cujo determinante é dado por

$$|\boldsymbol{\Sigma}_j| = \begin{cases} [\lambda_{x_j}\beta_j^2\sigma_{u_j}^2 + (\lambda_{x_j}+1)\sigma_{e_j}^2]\sigma_{u_j}^2 & \text{, se } \lambda_{x_j} \text{ é conhecido,} \\ [\lambda_{e_j}\sigma_{u_j}^2 + (\beta_j^2 + \lambda_{e_j})\sigma_{x_j}^2]\sigma_{u_j}^2 & \text{, se } \lambda_{e_j} \text{ é conhecido.} \end{cases} \tag{4.17}$$

### 4.2.1 A matriz de informação e parametrizações ortogonais

Seja $\mathbf{K} = \mathbf{K}(\boldsymbol{\theta})$ a matriz de informação do modelo definido em (4.4), onde $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \ldots, \boldsymbol{\theta}_k^\top)^\top$, com $\boldsymbol{\theta}_j$, $j = 1, \ldots, k$, definido em (4.15). Então da suposição de independência entre as populações, segue que a matriz de informação $\mathbf{K}$ é dada por

$$\mathbf{K} = \text{diag}(\mathbf{K}^{(1)}, \ldots, \mathbf{K}^{(k)}), \tag{4.18}$$

onde $\mathbf{K}^{(j)} = \mathbf{K}(\boldsymbol{\theta}_j)$, $j = 1, \ldots, k$, é a matriz de informação correspondente à $j$-ésima população. Arellano-Valle (1994) deriva uma expressão geral para a matriz de informação $\mathbf{K}(\boldsymbol{\theta}_j)$ e mostra que o fato desta matriz ser não diagonal dificulta, por exemplo, a determinação dos fatores de correção dos testes assintóticos da razão de verossimilhança e escore para hipóteses sobre os coeficientes angulares $\beta_1, \ldots, \beta_k$. Tais dificuldades podem ser removidas aplicando o procedimento de Cox e Reid (1987) para obter uma re-parametrização de modo que o parâmetro de interesse $\beta_j$ seja ortogonal aos outros parâmetros na população correspondente. Consideremos então a transformação de $\boldsymbol{\theta}_j$ (definido em (4.15)) por

$$\boldsymbol{\theta}_j = (\theta_{1j}, \theta_{2j}, \theta_{3j}, \theta_{4j}, \beta_j)^\top \to \boldsymbol{\phi}_j = (\phi_{1j}, \phi_{2j}, \phi_{3j}, \phi_{4j}, \beta_j)^\top,$$

de modo que $\theta_{lj} = \theta_{lj}(\boldsymbol{\phi}_j)$, $l = 1, 2, 3, 4$, sejam soluções das equações diferenciais

$$\sum_{l=1}^{4} K_{l,m}^{(j)} \frac{\partial \theta_{lj}}{\partial \beta_j} = -K_{\beta_j,m}^{(j)}, \quad l, m = 1, 2, 3, 4, \tag{4.19}$$

onde

$$K_{l,m}^{(j)} = n_j E\left[\frac{\partial L_j}{\partial \theta_{lj}} \frac{\partial L_j}{\partial \theta_{m_j}}\right] \quad \text{e} \quad K_{\beta_j,m}^{(j)} = n_j E\left[\frac{\partial L_j}{\partial \beta_j} \frac{\partial L_j}{\partial \theta_{m_j}}\right].$$

Uma solução de (4.19) é dada em Wong (1989) para o caso em que $\lambda_{e_j}$ é suposto conhecido, em Bolfarine e Cordani (1993) para o caso em que $\lambda_{x_j}$ (ou $k_{x_j}$) é assumido conhecido e ambos os casos unificados em Arellano-Valle (1994). Assim, uma solução de (4.19) segue ao escolher $\boldsymbol{\phi}_j = (\phi_{1j}, \phi_{2j}, \phi_{3j}, \phi_{4j}, \beta_j)^\top$ como

$$\begin{cases} \phi_{1j} = \alpha_j + \beta_j \mu_{x_j}, \quad \phi_{2j} = \mu_{x_j}, \quad \phi_{4j} = \sigma_{u_j}^2 \\ \phi_{3j} = \begin{cases} \lambda_{x_j} \beta_j^2 \sigma_{u_j}^2 + (\lambda_{x_j} + 1)\sigma_{e_j}^2 & \text{, se } \lambda_{x_j} \text{ é conhecido,} \\ (\beta_j^2 + \lambda_{e_j})\sigma_{x_j}^2 + \lambda_{e_j} \sigma_{u_j}^2 & \text{, se } \lambda_{e_j} \text{ é conhecido.} \end{cases} \end{cases} \tag{4.20}$$

Como remover $\beta_j$ de $\boldsymbol{\Sigma}_j$ não é simples, então, escolhendo $\phi_{1j} = E[Y_{ij}]$ e $\phi_{2j} = E[X_{ij}]$, temos de (4.20) que $\beta_j$ é removido de $\boldsymbol{\mu}_j$. Com relação aos parâmetros $\phi_{3j}$ e $\phi_{4j}$, estes são escolhidos de modo que $|\boldsymbol{\Sigma}_j| = \phi_{3j}\phi_{4j}$ (veja (4.17)) e que sejam ortogonais a $\beta_j$. Assim, temos que os parâmetros de locação $\phi_{1j}$ e $\phi_{2j}$ são globalmente ortogonais aos parâmetros de escala $\phi_{3j}$, $\phi_{4j}$ e $\beta_j$. Isto confirma o resultado de Mitchell (1989), que afirma que o parâmetro de locação $\boldsymbol{\mu}_j = \boldsymbol{\mu}(\phi_{1j}, \phi_{2j})$ e a matriz de dispersão $\boldsymbol{\Sigma}_j = \boldsymbol{\Sigma}(\phi_{3j}, \phi_{4j}, \beta_j)$ são globalmente ortogonais no modelo elíptico e, em particular, no modelo normal.

Seja $\boldsymbol{\phi}_{L_j} = (\phi_{1j}, \phi_{2j})^\top$ e $\boldsymbol{\phi}_{E_j} = (\phi_{3j}, \phi_{4j}, \beta_j)^\top$. Então $\boldsymbol{\mu}_j = E[\mathbf{Z}_{ij}]$ e $\boldsymbol{\Sigma}_j = \text{Var}[\mathbf{Z}_{ij}]$ são agora dadas por

$$\boldsymbol{\mu}_j = \boldsymbol{\mu}(\boldsymbol{\phi}_{L_j}) = (\phi_{1_j}, \phi_{2j})^\top \quad \text{e} \quad \boldsymbol{\Sigma}_j = \boldsymbol{\Sigma}(\boldsymbol{\phi}_{E_j}), \tag{4.21}$$

com

$$\boldsymbol{\Sigma}_j = \begin{cases} (\lambda_{x_j} + 1)^{-1} \begin{pmatrix} \phi_{3j} + (\lambda_{x_j}\beta_j)^2 \phi_{4j} & (\lambda_{x_j} + 1)\lambda_{x_j}\beta_j\phi_{4j} \\ (\lambda_{x_j} + 1)\lambda_{x_j}\beta_j\phi_{4j} & (\lambda_{x_j} + 1)^2 \phi_{4j} \end{pmatrix} & \text{, se } \lambda_{x_j} \text{ é conhecido,} \\ (\beta_j^2 + \lambda_{e_j})^{-1} \begin{pmatrix} \beta_j^2 \phi_{3j} + \lambda_{e_j}^2 \phi_{4j} & \beta_j(\phi_{3j} - \lambda_{e_j}\phi_{4j}) \\ \beta_j(\phi_{3j} - \lambda_{e_j}\phi_{4j}) & \phi_{3j} + \beta_j^2 \phi_{4j} \end{pmatrix} & \text{, se } \lambda_{e_j} \text{ é conhecido.} \end{cases} \tag{4.22}$$

Para facilitar a obtenção da matriz de informação, no que segue daremos algumas propriedades de $\mathbf{\Sigma}_j$, $j = 1, \ldots, k$, e suas derivadas, para ambas as parametrizações.

P1. $|\mathbf{\Sigma}_j| = \phi_{3j}\phi_{4j}$ ;

P2. $\mathbf{\Sigma}_j = \boldsymbol{\alpha}_{3j}\boldsymbol{\alpha}_{3j}^\top\phi_{3j} + \boldsymbol{\alpha}_{4j}\boldsymbol{\alpha}_{4j}^\top\phi_{4j}$,

onde

$$[\boldsymbol{\alpha}_{3j}, \boldsymbol{\alpha}_{4j}] = \begin{cases} \left\{(\lambda_{x_j}+1)^{-\frac{1}{2}}\binom{1}{0}, \ (\lambda_{x_j}+1)^{-\frac{1}{2}}\binom{\lambda_{x_j}\beta_j}{\lambda_{x_j}+1}\right\} & , \lambda_{x_j} \text{ é conhecido}, \\ \left\{(\beta_j^2+\lambda_{e_j})\binom{\beta_j}{1}, \ (\beta_j^2+\lambda_{e_j})\binom{\lambda_{e_j}}{-\beta_j}\right\} & , \lambda_{x_{e_j}} \text{ é conhecido}; \end{cases} \tag{4.23}$$

P3. $\mathbf{I}_2 = \overline{\boldsymbol{\alpha}}_{3j}\boldsymbol{\alpha}_{3j}^\top + \overline{\boldsymbol{\alpha}}_{4j}\boldsymbol{\alpha}_{4j}^\top$,

onde $\overline{\boldsymbol{\alpha}}_{lj} = \phi_{lj}\mathbf{\Sigma}_j^{-1}\boldsymbol{\alpha}_{lj}$, $l = 3, 4$, isto é,

$$[\overline{\boldsymbol{\alpha}}_{3j}, \overline{\boldsymbol{\alpha}}_{4j}] = \begin{cases} \left\{(\lambda_{x_j}+1)^{-\frac{1}{2}}\binom{\lambda_{x_j}+1}{-\lambda_{x_j}\beta_j}, \ (\lambda_{x_j}+1)^{-\frac{1}{2}}\binom{0}{1}\right\} & , \lambda_{x_j} \text{ é conhecido}, \\ \left\{(\beta_j^2+\lambda_{e_j})^{-\frac{1}{2}}\binom{\beta_j}{\lambda_{e_j}}, \ (\beta_j^2+\lambda_{e_j})^{-\frac{1}{2}}\binom{1}{-\beta_j}\right\} & , \lambda_{e_j} \text{ conhecido} \end{cases} \tag{4.24}$$

e

$$\overline{\boldsymbol{\alpha}}_{lj}^\top\boldsymbol{\alpha}_{lj} = 1, \quad l = 3, 4; \quad \overline{\boldsymbol{\alpha}}_{kj}^\top\boldsymbol{\alpha}_{lj} = 0, \quad k, l = 3, 4, \quad k \neq l;$$

P4. $\mathbf{\Sigma}_j^{-1} = \phi_{3j}^{-1}\overline{\boldsymbol{\alpha}}_{3j}\overline{\boldsymbol{\alpha}}_{3j}^\top + \phi_{4j}^{-1}\overline{\boldsymbol{\alpha}}_{4j}\overline{\boldsymbol{\alpha}}_{4j}^\top$;

P5. $\dfrac{\partial \mathbf{\Sigma}_j}{\partial \beta_j} = (\phi_{3j}\phi_{4j}/\sigma_{\beta_j}^2)^{\frac{1}{2}}(\boldsymbol{\alpha}_{3j}\boldsymbol{\alpha}_{4j}^\top + \boldsymbol{\alpha}_{4j}\boldsymbol{\alpha}_{3j}^\top)$,

onde

$$\sigma_{\beta_j}^2 = \begin{cases} \phi_{3j}/\lambda_{x_j}^2\phi_{4j} & , \text{para } \lambda_{x_j} \text{ conhecido}, \\ \left(\frac{\beta_j^2+\lambda_{e_j}}{\phi_{3j}-\lambda_{e_j}\phi_{4j}}\right)^2 \phi_{3j}\phi_{4j} & , \text{para } \lambda_{e_j} \text{ conhecido}. \end{cases} \tag{4.25}$$

Assim, sob a parametrização ortogonal e as propriedades dadas acima, segue de Arellano-Valle (1994) que a matriz de informação da $j$-ésima população é dada por

$$\mathbf{K}^{(j)} = \mathbf{K}(\boldsymbol{\phi}_j) = \text{diag}(\mathbf{K}_{L_j}, \mathbf{K}_{E_j}), \tag{4.26}$$

onde $\mathbf{K}_{L_j} = n_j\Sigma_j^{-1}$ e $\mathbf{K}_{E_j} = \text{diag}\{n_j/2\phi_{3j}^2, n_j/2\phi_{4j}^2, n_j/\sigma_{\beta_j}^2\}$ são as submatrizes de informação do vetor de locação $\boldsymbol{\phi}_{L_j} = (\phi_{1j}, \phi_{2j})^\top$ e escala $\boldsymbol{\phi}_{Ej} = (\phi_{3j}, \phi_{4j}, \beta_j)^\top$, respectivamente, e $\sigma_{\beta_j}^2 = \sigma_{\beta_j}^2(\boldsymbol{\phi}_{E_j})$ é como em (4.25).

### 4.2.2 Estimação por máxima verossimilhança usando a parametrização ortogonal

Seja $\widehat{\boldsymbol{\phi}} = (\widehat{\boldsymbol{\phi}}_1^\top, \ldots, \widehat{\boldsymbol{\phi}}_k^\top)^\top$ o EMV de $\boldsymbol{\phi} = (\boldsymbol{\phi}_1^\top, \ldots, \boldsymbol{\phi}_k^\top)^\top$, onde $\widehat{\boldsymbol{\phi}}_j$, $j = 1, \ldots, k$, é o EMV de $\boldsymbol{\phi}_j$ correspondente à $j$-ésima população. Então, usando a parametrização ortogonal definida em (4.20) e de (4.8) com $\boldsymbol{\theta}_j$ substituído por $\boldsymbol{\phi}_j$ segue que o EMV de $\boldsymbol{\phi}_j$, $j = 1, \ldots, k$, é solução das equações dadas por

$$\begin{aligned}
\frac{\partial L_j}{\partial \phi_{lj}}\bigg|_{\boldsymbol{\phi}_j = \widehat{\boldsymbol{\phi}}_j} &= \left[\sum_{i=1}^{n_j} \frac{\partial \boldsymbol{\mu}_j^\top}{\partial \phi_{lj}} \Sigma_j^{-1}(\mathbf{Z}_{ij} - \boldsymbol{\mu}_j)\right]_{\boldsymbol{\phi}_j = \widehat{\boldsymbol{\phi}_j}} = 0, \quad l = 1, 2, \\
\frac{\partial L_j}{\partial \phi_{lj}}\bigg|_{\boldsymbol{\phi}_j = \widehat{\boldsymbol{\phi}}_j} &= \left[-\frac{n_j}{2}\text{tr}\left(\Sigma_j^{-1}\frac{\partial \Sigma_j}{\partial \phi_{lj}}\right) + \frac{1}{2}\sum_{i=1}^{n_j}(\mathbf{Z}_{ij} - \boldsymbol{\mu}_j)^\top \Sigma_j^{-1}\frac{\partial \Sigma_j}{\partial \phi_{lj}}\Sigma_j^{-1}(\mathbf{Z}_{ij} - \boldsymbol{\mu}_j)\right]_{\boldsymbol{\phi}_j = \widehat{\boldsymbol{\phi}}_j} \\
&= 0, \quad l = 3, 4,
\end{aligned}$$

$$\frac{\partial L_j}{\partial \beta_j}\bigg|_{\boldsymbol{\phi}_j = \widehat{\boldsymbol{\phi}}_j} = \left[\frac{1}{2}\sum_{i=1}^{n_j}(\mathbf{Z}_{ij} - \boldsymbol{\mu}_j)^\top \Sigma_j^{-1}\frac{\partial \Sigma_j}{\partial \beta_j}\Sigma_j^{-1}(\mathbf{Z}_{ij} - \boldsymbol{\mu}_j)\right]_{\boldsymbol{\phi}_j = \widehat{\boldsymbol{\phi}}_j} = 0. \tag{4.27}$$

Da primeira equação, temos para cada $j = 1, \ldots, k$, que $\widehat{\boldsymbol{\mu}}_j = \mu(\widehat{\boldsymbol{\phi}}_{L_j}) = \overline{\mathbf{Z}}_j$ (definido em (4.9)), para ambas as condições de identificabilidade. Assim, o EMV dos parâmetros de locação $\phi_{1j}$ e $\phi_{2j}$ são dados por

$$\widehat{\phi}_{1j} = \overline{Y}_j \quad \text{e} \quad \widehat{\phi}_{2j} = \overline{X}_j, \tag{4.28}$$

onde

$$\overline{Y}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} Y_{ij} \quad \text{e} \quad \overline{X}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} X_{ij}.$$

Agora, considerando as propriedades da matriz $\mathbf{\Sigma}_j$ (veja as propriedades P1 - P5), temos, de (4.28) e da segunda e terceira equações em (4.27), que

$$\widehat{\phi}_{lj} = \overline{\boldsymbol{\alpha}}_{lj}^{\top}(\widehat{\beta}_j)\mathbf{S}_{n_j}\overline{\boldsymbol{\alpha}}_{lj}(\widehat{\beta}_j), \quad l = 3, 4, \tag{4.29}$$

e $\widehat{\beta}_j$ é raiz da equação

$$\overline{\boldsymbol{\alpha}}_{3j}^{\top}(\widehat{\beta}_j)\mathbf{S}_{n_j}\overline{\boldsymbol{\alpha}}_{4j}(\widehat{\beta}_j) = 0, \tag{4.30}$$

respectivamente, onde os vetores $\overline{\boldsymbol{\alpha}}_{lj}^{\top}(\widehat{\beta}_j)$, $l = 3, 4$, são como em (4.24), avaliados em $\widehat{\beta}_j$. Portanto, considerando (4.24), (4.29) e (4.30), segue que os EMV de $\phi_{3j}$, $\phi_{4j}$ e $\beta_j$ são dados a seguir

a) Para $\lambda_{x_j}$ conhecido,

$$\begin{aligned}
\widehat{\phi}_{3j} &= (\lambda_{x_j} + 1)S_{YY}^{(j)} - 2(\lambda_{x_j}\widehat{\beta}_j)S_{YX}^{(j)} + (\lambda_{x_j}\widehat{\beta}_j)^2\widehat{\phi}_{4j}, \\
\widehat{\phi}_{4j} &= \frac{S_{XX}^{(j)}}{\lambda_{x_j} + 1}, \\
\widehat{\beta}_j &= \left(\frac{\lambda_{x_j} + 1}{\lambda_{x_j}}\right)\frac{S_{YX}^{(j)}}{S_{XX}^{(j)}},
\end{aligned}$$

$j = 1, \ldots, k$. Substituindo $\widehat{\phi}_{4j}$ e $\widehat{\beta}_j$ em $\widehat{\phi}_{3j}$, obtemos

$$\widehat{\phi}_{3j} = (\lambda_{x_j} + 1)S_{YY.X}^{(j)},$$

onde $S_{YY.X}^{(j)} = S_{YY}^{(j)} - S_{YX}^{2(j)}/S_{XX}^{(j)}$, $j = 1, \ldots, k$;

b) Para $\lambda_{e_j}$ conhecido,

$$\begin{aligned}
\widehat{\phi}_{3j} &= \frac{\widehat{\beta}_j^2 S_{YX}^{(j)} + 2\lambda_{e_j}\widehat{\beta}_j^2 S_{YX}^{(j)} + \lambda_{e_j}S_{XX}^{(j)}}{\widehat{\beta}_j^2 + \lambda_{e_j}}, \\
\widehat{\phi}_{4j} &= \frac{S_{YY}^{(j)} - 2\widehat{\beta}_j S_{YX}^{(j)} + \widehat{\beta}_j^2 S_{XX}^{(j)}}{\widehat{\beta}_j^2 + \lambda_{e_j}}, \\
\widehat{\beta}_j &= \frac{S_{YY}^{(j)} - \lambda_{e_j}S_{XX}^{(j)} + \{(S_{YY}^{(j)} - \lambda_{e_j}S_{XX}^{(j)})^2 + 4\lambda_{e_j}S_{YX}^{(j)2}\}^{1/2}}{2S_{YX}^{(j)}},
\end{aligned}$$

$j = 1, \ldots, k$.

Notemos que em ambos os casos de identificabilidade

$$\widehat{\phi}_{3j}\widehat{\phi}_{4j} = |\Sigma(\widehat{\boldsymbol{\phi}}_{E_j})| = S^{(j)}_{YY.X} S^{(j)}_{XX}, \tag{4.31}$$

$j = 1, \ldots, k$.

### 4.2.3 Propriedades do EMV

Para estudar as propriedades de $\widehat{\boldsymbol{\phi}}_j = (\widehat{\boldsymbol{\phi}}^{\top}_{L_j}, \widehat{\boldsymbol{\phi}}^{\top}_{E_j})^{\top}$ temos, de (4.13) e da suposição de normalidade, que

$$\widehat{\boldsymbol{\mu}}_j \sim N_2\left(\boldsymbol{\mu}_j, \frac{1}{n_j}\boldsymbol{\Sigma}_j\right) \quad \text{e} \quad \widehat{\boldsymbol{\Sigma}}_j \sim W_2\left(n_j - 1, \frac{1}{n_j}\boldsymbol{\Sigma}_j\right), \tag{4.32}$$

são independentes, onde $\boldsymbol{\mu}_j$ e $\boldsymbol{\Sigma}_j$ são como em (4.5) e $W_k(m, \mathbf{A})$ denota a distribuição de Wishart $k$-variada com matriz de dispersão $\mathbf{A}$ e $m$ graus de liberdade. Sob a parametrização ortogonal temos que

$$\widehat{\boldsymbol{\phi}}_{L_j} = \phi_L(\widehat{\boldsymbol{\mu}}_j) = (\widehat{\phi}_{1j}, \widehat{\phi}_{2j})^{\top} \quad \text{e} \quad \widehat{\boldsymbol{\phi}}_{E_j} = \phi_E(\widehat{\boldsymbol{\Sigma}}_j) = (\widehat{\phi}_{3j}, \widehat{\phi}_{4j}, \widehat{\beta}_j)^{\top}$$

com $\widehat{\boldsymbol{\Sigma}}_j = \mathbf{S}_{n_j}$, de onde segue que

i) $\widehat{\boldsymbol{\phi}}_{L_j} = \overline{\mathbf{Z}}_j \sim N_2(\boldsymbol{\phi}_{L_j}, \frac{1}{n_j}\boldsymbol{\Sigma}_j)$ e ii) $\widehat{\boldsymbol{\phi}}_{L_j}$ e $\widehat{\boldsymbol{\phi}}_{E_j}$ são independentes.

Seja $\boldsymbol{\phi} = (\boldsymbol{\phi}^{\top}_L, \boldsymbol{\phi}^{\top}_E)^{\top}$, onde $\boldsymbol{\phi}_L = (\boldsymbol{\phi}^{\top}_{L_1}, \ldots, \boldsymbol{\phi}^{\top}_{L_k})^{\top}$ e $\boldsymbol{\phi}_E = (\boldsymbol{\phi}^{\top}_{E_1}, \ldots, \boldsymbol{\phi}^{\top}_{E_k})^{\top}$. Então $\widehat{\boldsymbol{\phi}}_L = (\widehat{\boldsymbol{\phi}}^{\top}_{L_1}, \ldots, \widehat{\boldsymbol{\phi}}^{\top}_{L_k})^{\top}$ e $\widehat{\boldsymbol{\phi}}_E = (\widehat{\boldsymbol{\phi}}^{\top}_{E_1}, \ldots, \widehat{\boldsymbol{\phi}}^{\top}_{E_k})^{\top}$ são independentes e $\widehat{\boldsymbol{\phi}}_L \sim N_{2k}(\boldsymbol{\phi}_L, \boldsymbol{\Sigma}^{(*)})$, onde $\boldsymbol{\Sigma}^{(*)} = \text{Diag}(\frac{1}{n_1}\boldsymbol{\Sigma}_1, \ldots, \frac{1}{n_k}\boldsymbol{\Sigma}_k)$. Agora, para cada $j = 1, \ldots, k$, a distribuição de $\widehat{\boldsymbol{\phi}}_{E_j}$ depende da distribuição de $\mathbf{S}_{n_j}$ e, conseqüentemente, as distribuições de $\widehat{\phi}_{3j}$, $\widehat{\phi}_{4j}$ e $\widehat{\beta}_j$ dependem da distribuição $\mathbf{S}_{n_j}$ para ambos os casos de identificabilidade. Como, sob normalidade do modelo $\mathbf{S}_{n_j} \sim W_2(n_j - 1, \frac{1}{n_j}\boldsymbol{\Sigma}_j)$, tem-se que (veja Muirhead, 1982)

i) $S^{(j)}_{YY.X} = S^{(j)}_{YY} - S^{2(j)}_{XY}/S^{(j)}_{XX}$ é independente de $(S^{(j)}_{YX}, S^{(j)}_{XX})^{\top}$,

ii) $n_j S_{YY.X}^{(j)} / \sigma_{YY.X}^{(j)} \sim \chi^2_{n_j-2}$,

onde $\sigma_{YY.X}^{(j)} = \sigma_{YY}^{(j)} - \sigma_{XY}^{2(j)}/\sigma_{XX}^{(j)}$,

iii) $(S_{YX}^{(j)}/S_{XX}^{(j)}) \sim N\left(\dfrac{\sigma_{YX}^{(j)}}{\sigma_{XX}^{(j)}} S_{XX}^{(j)}, \dfrac{\sigma_{YY.X}^{(j)}}{n_j} S_{XX}^{(j)}\right)$,

iv) $n_j \dfrac{S_{YY}^{(j)}}{\sigma_{YY}^{(j)}} \sim \chi^2_{n_j-1}$ e $n_j \dfrac{S_{XX}^{(j)}}{\sigma_{XX}^{(j)}} \sim \chi^2_{n_j-1}$.

Note que sob a condição de identificabilidade $\lambda_{x_j}$, conhecido, temos que

$$\phi_{3j} = (\lambda_{x_j} + 1)\sigma_{YY.X}^{(j)}, \quad \phi_{4j} = \frac{\sigma_{XX}^{(j)}}{\lambda_{x_j} + 1} \quad \text{e} \quad \beta_j = \left(\frac{\lambda_{x_j} + 1}{\lambda_{x_j}}\right)\frac{\sigma_{XY}^{(j)}}{\sigma_{XX}^{(j)}}$$

com EMV dados por

$$\begin{aligned}
\widehat{\phi}_{3j} &= (\lambda_{x_j} + 1) S_{YY.X}^{(j)}, \\
\widehat{\phi}_{4j} &= \frac{S_{XX}^{(j)}}{\lambda_{x_j} + 1}, \\
\widehat{\beta}_j &= \left(\frac{\lambda_{x_j} + 1}{\lambda_{x_j}}\right)\frac{S_{XY}^{(j)}}{S_{XX}^{(j)}},
\end{aligned}$$

$j = 1, \ldots, k$, respectivamente. Assim, usando as propriedades i) – iv), temos que (veja Arellano-Valle, 1994)

1) $\widehat{\phi}_{3j}$ e $\widehat{\phi}_{4j}$ são independentes;

2) $n_j \dfrac{\widehat{\phi}_{3j}}{\phi_{3j}} \sim \chi^2_{n_j-2}$ e $n_j \dfrac{\widehat{\phi}_{4j}}{\phi_{4j}} \sim \chi^2_{n_j-1}$;

3) $\widehat{\phi}_{4j}^{\frac{1}{2}}(\widehat{\beta}_j - \beta_j) \sim N\left(0, \dfrac{\sigma^2_{\beta_j}\phi_{4j}}{n_j}\right)$, onde $\sigma^2_{\beta_j} = \frac{\phi_{3j}}{\lambda_{x_j}\phi_{4j}}$;

4) $\widehat{\beta}_j \sim t\left(\beta_j, \dfrac{\sigma^2_{\beta_j}}{n_j - 1}; n_j - 2\right)$,

$j = 1, \ldots, k$.

## 4.3 O modelo estrutural elíptico com mais de uma população

### 4.3.1 Introdução

Os modelo elípticos, dos quais o modelo normal é um membro particular, têm sido motivo de numerosas pesquisas nestes últimos anos. Por exemplo, no contexto dos modelos lineares, vários autores têm usado estes modelos para explorar as conseqüências de se considerar os erros do modelo tendo distribuição com caudas mais pesadas que a distribuição normal, como por exemplo, a distribuição $t$-Student. Contudo, como visto no capítulo 2, aparecem duas formas de especificar o modelo estrutural elíptico quando se tem uma população:

i) **Modelo elíptico dependente:** assume que os erros têm distribuição conjunta elíptica, de modo que eles podem ter covariâncias nulas sem serem independentes. Para este modelo, veja por exemplo, Zellner (1976), Ghosh e Sinha (1980), Anderson et al. (1986);

ii) **Modelo elíptico independente:** supõe que os erros (ou as observações) são independentes e têm distribuição marginal elíptica. Para este modelo veja, por exemplo, Lang, Little e Taylor (1984), Taylor (1992), Kano, Berkane e Bentler (1993) e Arellano-Valle (1994).

Com o intuito de estender o modelo definido em (4.5) a uma versão elíptica, vamos considerar para tais extensões as especificações para o modelo elíptico quando se tem uma população, isto é, as extensões vão depender da distribuição conjunta dos $x_{ij}$, $e_{ij}$ e $u_{ij}$, como veremos na seção seguinte.

### 4.3.2 Especificação do modelo estrutural elíptico para mais de uma população

Note que o modelo em (4.1) pode ser escrito como

$$\mathbf{Z}_{ij} = \mathbf{a}_j + \mathbf{B}_j\mathbf{r}_{ij}, \tag{4.33}$$

$j = 1, \ldots, k$, $i = 1, \ldots, n_j$, então, similarmente ao modelo normal, a especificação da distribuição de $\mathbf{Z}_{ij}$ depende da especificação da distribuição de $\mathbf{r}_{ij}$. Neste trabalho vamos supor que para cada $j = 1, \ldots, k$, os $\mathbf{r}_{1j}, \ldots, \mathbf{r}_{n_j j}$ são não correlacionados e que $\mathbf{r}_{ij} \sim E\ell_3(\boldsymbol{\eta}_j, \boldsymbol{\Omega}_j; \phi)$, $i = 1, \ldots, n_j$, onde

$$\boldsymbol{\eta}_j = \begin{pmatrix} \mu_{x_j} \\ 0 \\ 0 \end{pmatrix} \quad \text{e} \quad \boldsymbol{\Omega}_j = \begin{pmatrix} \sigma^2_{x_j} & 0 & 0 \\ 0 & \sigma^2_{e_j} & 0 \\ 0 & 0 & \sigma^2_{u_j} \end{pmatrix}. \tag{4.34}$$

Além disso, os $\mathbf{r}_{ij}$ correspondentes às observações de populações distintas são não correlacionados, isto é, $\text{Cov}[\mathbf{r}_{ij}, \mathbf{r}_{lm}] = 0$, $j, m = 1, \ldots, k$, $j \neq m$, $i = 1, \ldots, n_j$, $l = 1, \ldots, n_m$.

Seja

$$\mathbf{r}_{(j)} = \begin{pmatrix} \mathbf{r}_{1j} \\ \mathbf{r}_{2j} \\ \vdots \\ \mathbf{r}_{n_j j} \end{pmatrix}, \quad j = 1, \ldots, k, \quad \text{e} \quad \mathbf{r} = \begin{pmatrix} \mathbf{r}_{(1)} \\ \mathbf{r}_{(2)} \\ \vdots \\ \mathbf{r}_{(k)} \end{pmatrix}. \tag{4.35}$$

Então, vamos considerar três possíveis extensões do modelo sob normalidade para modelos elípticos, a saber,

(A) $\mathbf{r} \sim E\ell_{3n}(\boldsymbol{\eta}, \boldsymbol{\Omega}; \phi)$;

(B) $\mathbf{r} \sim E\ell^{(I)}_{3n}(\boldsymbol{\eta}, \boldsymbol{\Omega}; \phi) \Leftrightarrow \mathbf{r}_{(j)} \overset{\text{ind}}{\sim} E\ell_{3n_j}(\boldsymbol{\eta}_{(j)}, \boldsymbol{\Omega}_{(j)}; \phi)$, $j = 1, \ldots, k$;

(C) $\mathbf{r} \sim E\ell^{(II)}_{3n}(\boldsymbol{\eta}, \boldsymbol{\Omega}; \phi) \Leftrightarrow \mathbf{r}_{(j)} \sim E\ell^{(I)}_{3n_j}(\boldsymbol{\eta}_{(j)}, \boldsymbol{\Omega}_{(j)}; \phi)$, $j = 1, \ldots, k$,

onde

$$\boldsymbol{\eta} = \begin{pmatrix} \boldsymbol{\eta}_{(1)} \\ \boldsymbol{\eta}_{(2)} \\ \vdots \\ \boldsymbol{\eta}_{(k)} \end{pmatrix}, \quad \boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_{(1)} & \cdots & 0 \\ \vdots & \boldsymbol{\Omega}_{(2)} & \vdots \\ 0 & \cdots & \boldsymbol{\Omega}_{(k)} \end{pmatrix}, \tag{4.36}$$

$$\boldsymbol{\eta}_{(j)} = \mathbf{1}_{n_j} \otimes \boldsymbol{\eta}_j \quad \text{e} \quad \boldsymbol{\Omega}_{(j)} = \mathbf{I}_{n_j} \otimes \boldsymbol{\Omega}_j, \tag{4.37}$$

com $\boldsymbol{\eta}_j$ e $\boldsymbol{\Omega}_j$ como em (4.34).

Em qualquer das três especificações temos que, desde que existam,

$$E[\mathbf{r}] = \boldsymbol{\eta} \quad \text{e} \quad \mathrm{Var}[\mathbf{r}] = \boldsymbol{\alpha}_\phi \boldsymbol{\Omega}, \tag{4.38}$$

isto é,

$$E[\mathbf{r}_{(j)}] = \boldsymbol{\eta}_{(j)}, \quad \mathrm{Var}[\mathbf{r}_{(j)}] = \boldsymbol{\alpha}_\phi \boldsymbol{\Omega}_{(j)} \quad \text{e} \quad \mathrm{Cov}[\mathbf{r}_{(j)}, \mathbf{r}_{(j')}] = 0,$$

$j, j' = 1, \ldots, k$ e $j \neq j'$, onde $\boldsymbol{\alpha}_\phi = -2\phi'(0)$.

Em (A) os vetores $\mathbf{r}_{(1)}, \ldots, \mathbf{r}_{(k)}$ não são independentes. Esta mesma propriedade tem a seqüência $\mathbf{r}_{1j}, \ldots, \mathbf{r}_{n_j j}$, para cada $j = 1, \ldots, k$, a menos que se tenha normalidade. Na especificação (B), $\mathbf{r}_{(1)}, \ldots, \mathbf{r}_{(k)}$ são independentes, mas a seqüência $\mathbf{r}_{1j}, \ldots, \mathbf{r}_{n_j j}$ não é independente. Finalmente, na especificação (C), a seqüência $\mathbf{r}_{1j}, \ldots, \mathbf{r}_{n_j j}$ é independente para cada $j = 1, \ldots, k$, e tem distribuição comum $\mathbf{r}_{ij} \sim E\ell_3(\boldsymbol{\eta}_j, \boldsymbol{\Omega}_j; \phi)$, de modo que a distribuição de $\mathbf{r}_{(j)}$ não é elíptica.

Sob normalidade as três especificações coincidem. Por outro lado, se $k = 1$ (caso de uma população), então a primeira e segunda especificações coincidem com o modelo elíptico dependente e a terceira especificação coincide com o modelo elíptico independente definido em Arellano-Valle (1994). Note que a amostra de tamanho $n_j$ da $j$-ésima população pode ser colocada na forma $\mathbf{Z}_{(j)} = (\mathbf{Z}_{1j}^\top, \ldots, \mathbf{Z}_{n_j j}^\top)^\top$. Assim, de (4.33) temos que

$$\mathbf{Z}_{(j)} = \mathbf{1}_{n_j} \otimes \mathbf{a}_j + (\mathbf{I}_{n_j} \otimes \mathbf{B}_j)\mathbf{r}_{(j)}, j = 1, \ldots, k, \tag{4.39}$$

onde $\mathbf{r}_{(j)}$ é como em (4.35). Além disso, as observações de tamanho $n = n_1 + \cdots + n_k$ do modelo em (4.33) pode ser colocado no vetor $\mathbf{Z} = (\mathbf{Z}_{(1)}^\top, \ldots, \mathbf{Z}_{(k)}^\top)^\top$, que por (4.39) pode ser expresso como

$$\mathbf{Z} = \mathbf{a} + \mathbf{B}\mathbf{r}, \tag{4.40}$$

onde

$$\mathbf{a} = \begin{pmatrix} \mathbf{1}_{n_1} \otimes \mathbf{a}_1 \\ \vdots \\ \mathbf{1}_{n_k} \otimes \mathbf{a}_k \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} \mathbf{I}_{n_1} \otimes \mathbf{B}_1 & \cdots & 0 \\ \vdots & \mathbf{I}_{n_2} \otimes \mathbf{B}_2 & \vdots \\ 0 & \cdots & \mathbf{I}_{n_k} \otimes \mathbf{B}_k \end{pmatrix} \tag{4.41}$$

e $\mathbf{r}$ é como em (4.35).

Assim, aplicando o Teorema A.3 do Apêndice A, em (4.40), (4.39) e (4.33) temos, sob as especificações dos modelos elípticos, que

(A) $\mathbf{Z} \sim E\ell_{2n}(\boldsymbol{\mu}_{(*)}, \boldsymbol{\Sigma}_{(*)}; \phi)$ ,

(B) $\mathbf{Z} \sim E\ell_{2n}^{(I)}(\boldsymbol{\mu}_{(*)}, \boldsymbol{\Sigma}_{(*)}; \phi) \Leftrightarrow \mathbf{Z}_{(j)} \overset{\text{ind}}{\sim} E\ell_{2n_j}(\boldsymbol{\mu}_{(j)}, \boldsymbol{\Sigma}_{(j)}; \phi)$ ,

(C) $\mathbf{Z} \sim E\ell_{2n}^{(II)}(\boldsymbol{\mu}_{(*)}, \boldsymbol{\Sigma}_{(*)}; \phi) \Leftrightarrow \mathbf{Z}_{(j)} \sim E\ell_{2n_j}^{(I)}(\boldsymbol{\mu}_{(j)}, \boldsymbol{\Sigma}_{(j)}; \phi)$ ,

onde

$$\boldsymbol{\mu}_{(*)} = \begin{pmatrix} \boldsymbol{\mu}_{(1)} \\ \boldsymbol{\mu}_{(2)} \\ \vdots \\ \boldsymbol{\mu}_{(k)} \end{pmatrix} \quad \text{e} \quad \boldsymbol{\Sigma}_{(*)} = \begin{pmatrix} \boldsymbol{\Sigma}_{(1)} & \cdots & 0 \\ \vdots & \boldsymbol{\Sigma}_{(2)} & \vdots \\ 0 & \cdots & \boldsymbol{\Sigma}_{(k)} \end{pmatrix} \tag{4.42}$$

com

$$\boldsymbol{\mu}_{(j)} = \mathbf{1}_{n_j} \otimes \boldsymbol{\mu}_j, \quad \boldsymbol{\Sigma}_{(j)} = \mathbf{I}_{n_j} \otimes \boldsymbol{\Sigma}_j, \quad \boldsymbol{\mu}_j = \mathbf{a}_j + \mathbf{B}_j\boldsymbol{\eta}_j \quad \text{e} \quad \boldsymbol{\Sigma}_j = \mathbf{B}_j\boldsymbol{\Omega}_j\mathbf{B}_j^{\top},$$

$j = 1, \ldots, k$ ($\boldsymbol{\eta}_j$ e $\boldsymbol{\Omega}_j$ são como em (4.34) e $\boldsymbol{\mu}_j$ e $\boldsymbol{\Sigma}_j$ como em (4.5)).

Note que as inversas das matrizes $\boldsymbol{\Sigma}_{(j)}$ e $\boldsymbol{\Sigma}$ são dadas por

$$\boldsymbol{\Sigma}_{(j)}^{-1} = \mathbf{I}_{n_j} \otimes \boldsymbol{\Sigma}_j^{-1} \quad \text{e} \quad \boldsymbol{\Sigma}^{-1} = \text{Diag}(\boldsymbol{\Sigma}_{(1)}^{-1}, \ldots, \boldsymbol{\Sigma}_{(k)}^{-1}), \tag{4.43}$$

respectivamente. Assim, os vetores aleatórios

$$\mathbf{T} = \boldsymbol{\Sigma}_{(*)}^{-1/2}(\mathbf{Z} - \boldsymbol{\mu}_{(*)}), \quad \mathbf{T}_{(j)} = \boldsymbol{\Sigma}_{(j)}^{-1/2}(\mathbf{Z}_{(j)} - \boldsymbol{\mu}_{(j)}) \quad \text{e} \quad \mathbf{T}_{ij} = \boldsymbol{\Sigma}_j^{-1/2}(\mathbf{Z}_{ij} - \boldsymbol{\mu}_j) \tag{4.44}$$

têm uma distribuição esférica, isto é, sob as especificações dos modelos definidos em (A), (B) e (C), temos que

(A) $\mathbf{T} \sim E\ell_{2n}(\mathbf{0}, \mathbf{I}_{2n}; \phi), \quad n = n_1 + \cdots + n_k,$

(B) $\mathbf{T} \sim E\ell_{2n}^{(I)}(\mathbf{0}, \mathbf{I}_{2n}; \phi) \Leftrightarrow \mathbf{T}_{(j)} \overset{\text{ind}}{\sim} E\ell_{2n_j}(\mathbf{0}, \mathbf{I}_{2n_j}; \phi)$, $j = 1, \ldots, k$,

(C) $\mathbf{T} \sim E\ell_{2n}^{(II)}(\mathbf{0}, \mathbf{I}_{2n}; \phi) \Leftrightarrow \mathbf{T}_{(j)} \sim E\ell_{2n_j}^{(I)}(\mathbf{0}, \mathbf{I}_{2n_j}; \phi)$, $j = 1, \ldots, k$,

respectivamente.

Note também que, desde que existam,

$$E[\mathbf{Z}] = \boldsymbol{\mu}_{(*)} \quad \text{e} \quad \text{Var}[\mathbf{Z}] = \boldsymbol{\alpha}_{\phi}\Sigma_{(*)}, \tag{4.45}$$

isto é,

$$E[\mathbf{Z}_{(j)}] = \boldsymbol{\mu}_{(j)}, \quad \text{Var}[\mathbf{Z}_{(j)}] = \boldsymbol{\alpha}_{\phi}\boldsymbol{\Sigma}_{(j)} \quad \text{e} \quad \text{Cov}[\mathbf{Z}_{(j)}, \mathbf{Z}_{(j')}] = 0, \tag{4.46}$$

$j, j' = 1, \ldots, k$, $j \neq j'$.

Similarmente ao modelo normal, temos que para cada $j = 1, \ldots, k$,

$$\boldsymbol{\mu}_j = \boldsymbol{\mu}(\boldsymbol{\theta}_j) \quad \text{e} \quad \boldsymbol{\Sigma}_j = \Sigma(\boldsymbol{\theta}_j), \tag{4.47}$$

com $\boldsymbol{\theta}_j = (\alpha_j, \mu_{x_j}, \sigma_{x_j}^2, \sigma_{e_j}^2, \sigma_{u_j}^2, \beta_j)^\top$. Sendo $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \ldots, \boldsymbol{\theta}_k^\top)^\top$ o vetor de parâmetros e sob a hipótese adicional de existência de densidade do vetor $\mathbf{Z} = (\mathbf{Z}_{(1)}^\top, \ldots, \mathbf{Z}_{(k)}^\top)^\top$, temos, sob as especificações (A), (B) e (C), as seguintes densidades conjuntas:

$$\begin{aligned} (\text{A}): \qquad p_{\mathbf{Z}}(\mathbf{Z}|\boldsymbol{\theta}) &= |\Sigma_{(*)}|^{-1/2} f((\mathbf{Z} - \boldsymbol{\mu}_{(*)})^\top \Sigma_{(*)}^{-1}(\mathbf{Z} - \boldsymbol{\mu}_{(*)}) \\ &= \prod_{i=1}^{k} |\Sigma_j|^{-n_j/2} f\left(\sum_{j=1}^{k}\sum_{i=1}^{n_j}(\mathbf{Z}_{ij} - \boldsymbol{\mu}_j)\Sigma_j^{-1}(\mathbf{Z}_{ij} - \boldsymbol{\mu}_j)\right), \end{aligned} \tag{4.48}$$

onde $f(\cdot)$ é uma densidade esférica em $\mathbb{R}^{2n}$, com $n = n_1 + n_2 + \cdots + n_k$;

$$\begin{aligned} (\text{B}): \qquad p_{\mathbf{Z}}(\mathbf{Z}|\boldsymbol{\theta}) &= \prod_{j=1}^{k} P_{\mathbf{Z}_{(j)}}(\mathbf{Z}_{(j)}/\boldsymbol{\theta}_j) \\ &= \prod_{j=1}^{k} |\Sigma_j|^{-n_j/2} g\left(\sum_{i=1}^{n_j}(\mathbf{Z}_{ij} - \boldsymbol{\mu}_j)\Sigma_j^{-1}(\mathbf{Z}_{ij} - \boldsymbol{\mu}_j)\right), \end{aligned} \tag{4.49}$$

onde $g(\cdot)$ é uma densidade esférica em $\mathbb{R}^{2n_j}$.

$$\begin{aligned} (\text{C}): \qquad p_{\mathbf{Z}}(\mathbf{Z}|\boldsymbol{\theta}) &= \prod_{j=1}^{k}\prod_{i=1}^{n_j} |\Sigma_j|^{-1/2} h((\mathbf{Z}_{ij} - \boldsymbol{\mu}_j)^\top \Sigma_j^{-1}(\mathbf{Z}_{ij} - \boldsymbol{\mu}_j)) \\ &= \left(\prod_{j=1}^{k} |\Sigma_j|^{-n_j/2}\right) \prod_{j=1}^{k}\prod_{i=1}^{n_j} h((\mathbf{Z}_{ij} - \boldsymbol{\mu}_j)^\top \Sigma_j^{-1}(\mathbf{Z}_{ij} - \boldsymbol{\mu}_j)), \end{aligned} \tag{4.50}$$

onde $h(\cdot)$ é uma densidade esférica em $\mathbb{R}^2$. Em qualquer das especificações acima, as funções $f$, $g$ e $h$ não dependem do parâmetro $\boldsymbol{\theta}$.

### 4.3.3 Identificabilidade no modelo estrutural elíptico

Sob normalidade, Wong (1991) e Nascimento (1994) provaram que o modelo (4.33) é identificável sob as condições $\lambda_{e_j}$ ou $\lambda_{x_j}$, conhecido, $j = 1, \ldots, k$. Como no caso elíptico, a função de verossimilhança para uma amostra de tamanho $n = n_1 + \cdots + n_k$ é dada por (4.48), (4.49) ou (4.50), dependendo do modelo que será objeto de estudo. Assim, a distribuição de $\mathbf{Z}$ depende de $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \ldots, \boldsymbol{\theta}_k^\top)^\top$ através de $\boldsymbol{\mu}(\boldsymbol{\theta}_1), \ldots, \boldsymbol{\mu}(\boldsymbol{\theta}_k)$, $\Sigma(\boldsymbol{\theta}_1), \ldots, \Sigma(\boldsymbol{\theta}_k)$, com $\boldsymbol{\theta}_j = (\alpha_j, \mu_{x_j}, \sigma_{x_j}^2, \sigma_{e_j}^2, \sigma_{u_j}^2, \beta_j)^\top$. Como as observações são não correlacionadas através das populações, então a identificabilidade do modelo elíptico sob qualquer das três especificações depende da identificabilidade em cada uma das $k$ populações, que corresponde a cada um dos $k$ modelos estruturais elípticos serem não identificáveis (veja Arellano-Valle, 1994, p.190). Assim, segue que qualquer dos modelos estruturais elípticos sob as especificações (A), (B) ou (C) é não identificável, pois, por exemplo, se $\boldsymbol{\theta}_j = (1, 1, 1, 1, 1, 1)^\top$ e $\boldsymbol{\theta}_j^* = (3/2, 1, 2, 3/2, 0, 1)^\top$, $j = 1, \ldots, k$, então temos que

$$\boldsymbol{\mu}_{(*)}(\boldsymbol{\theta}) = \boldsymbol{\mu}_{(*)}(\boldsymbol{\theta}^*) \quad \text{e} \quad \Sigma_{(*)}(\boldsymbol{\theta}) = \Sigma_{(*)}(\boldsymbol{\theta}^*),$$

dado que $\boldsymbol{\mu}(\boldsymbol{\theta}_j) = \boldsymbol{\mu}(\boldsymbol{\theta}_j^*) = \binom{2}{1}$ e $\Sigma(\boldsymbol{\theta}_j) = \Sigma(\boldsymbol{\theta}_j^*) = \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}$, para todo $j = 1, \ldots, k$.

Contudo, sob as condições $\lambda_{e_j}$ ou $\lambda_{x_j}$, conhecidos, $j = 1, \ldots, k$, temos que os modelos sob as especificações (A), (B) ou (C) são identificáveis pois sob estas condições de identificabilidade $\boldsymbol{\theta}_j$ fica com um total de cinco parâmetros (como em (4.15)) e, logo, sob qualquer das três especificações o modelo em estudo contém $5k$ parâmetros. Por outro lado, como

$$\boldsymbol{\mu}_j = \begin{pmatrix} \mu_Y^{(j)} \\ \mu_X^{(j)} \end{pmatrix} \quad \text{e} \quad \Sigma_j = \Sigma(\boldsymbol{\theta}_j) = \begin{pmatrix} \sigma_{YY}^{(j)} & \sigma_{XY}^{(j)} \\ \sigma_{XY}^{(j)} & \sigma_{XX}^{(j)} \end{pmatrix}, \tag{4.51}$$

$j = 1, \ldots, k$, temos que

$$(\boldsymbol{\theta}_1^\top, \ldots, \boldsymbol{\theta}_k^\top)^\top \to (\mu(\boldsymbol{\theta}_1), \Sigma(\boldsymbol{\theta}_1), \ldots, \mu(\boldsymbol{\theta}_k), \Sigma(\boldsymbol{\theta}_k))$$

é um a um, e esta afirmação garante a identificabilidade do modelo estrutural elíptico sob qualquer das três especificações.

Sob qualquer das duas condições de identificabilidade temos sob o modelo elíptico com a especificação (A), com $f$ decrescente em $(0, +\infty)$, temos que os EMV de $\boldsymbol{\mu}_{(*)}$ e $\boldsymbol{\Sigma}_{(*)}$ são dados por

$$\widehat{\boldsymbol{\mu}}_{(*)} = \widehat{\boldsymbol{\mu}}_{(*)N} \quad \text{e} \quad \widehat{\boldsymbol{\Sigma}}_{(*)} = \frac{np}{u_f}\widehat{\boldsymbol{\Sigma}}_{(*)_N}, \quad n = n_1 + \cdots + n_k, \tag{4.52}$$

respectivamente, onde $\widehat{\boldsymbol{\mu}}_{(*)N}$ e $\widehat{\boldsymbol{\Sigma}}_{(*)N}$ são os respectivos EMV sob normalidade, $p = 2$ (dimensão de $\mathbf{Z}_{ij}$ e $p < n_j$) e $u_f$ é o máximo da função $u^{np/2}f(u)$. Por exemplo, se $f(u) = k(p, \nu)\{\nu + u\}^{-(\nu+p)/2}$ que corresponde ao modelo $t$-Student com $\nu$ graus de liberdade, é fácil ver que $u_f = np$, para todo $\nu > 0$. O resultado acima é conseqüência do Teorema 1 dado em Anderson, Fang e Hsu (1986) (veja também o capítulo 2, seção 2.6). Assim, de (4.52) temos para $j = 1, \ldots, k$, que

$$\widehat{\boldsymbol{\mu}}_j = \overline{\mathbf{Z}}_j, \quad \widehat{\boldsymbol{\Sigma}}_j = \frac{np}{u_f}\mathbf{S}_{n_j}. \tag{4.53}$$

Similarmente, sob o modelo elíptico com a especificação (B), com $g$ decrescente em $(0, +\infty)$, temos que os EMVs de $\boldsymbol{\mu}_j$ e $\boldsymbol{\Sigma}_j$, são obtidos a partir das equações

$$\widehat{\boldsymbol{\mu}}_{(j)} = \widehat{\boldsymbol{\mu}}_{N(j)} \quad \text{e} \quad \widehat{\boldsymbol{\Sigma}}_{(j)} = \frac{n_j p}{u_g}\widehat{\boldsymbol{\Sigma}}_{N(j)}, \tag{4.54}$$

$j = 1, \ldots, k$. Assim, o EMV de $\boldsymbol{\mu}_j$ e $\boldsymbol{\Sigma}_j$ são dados por

$$\widehat{\mu}_j = \overline{\mathbf{Z}}_j \quad \text{e} \quad \widehat{\boldsymbol{\Sigma}}_j = \frac{n_j p}{u_g}\mathbf{S}_{n_j}, \tag{4.55}$$

$j = 1, \ldots, k$. No modelo elíptico segundo a especificação (C), os EMV dos parâmetros não têm uma forma fechada. Neste caso, sugere-se usar algum método de aproximação.

### 4.3.4 Matriz de informação e parametrização ortogonal

Seja $L = L(\boldsymbol{\theta})$, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \ldots, \boldsymbol{\theta}_n^\top)^\top$ o logaritmo da função de verossimilhança do modelo estrutural elíptico definido nas especificações da seção anterior. Então,

$$L = \begin{cases} \sum_{j=1}^{k} \left(-\frac{n_j}{2}\right) \log |\Sigma_j| + \log f(\|\mathbf{T}\|^2), & \text{no caso (A)}, \\ \sum_{j=1}^{k} \left(-\frac{n_j}{2}\right) \log |\Sigma_j| + \sum_{j=1}^{n_j} \log g(\|\mathbf{T}_{(j)}\|^2), & \text{no caso (B)}, \\ \sum_{j=1}^{k} \left(-\frac{n_j}{2}\right) \log |\Sigma_j| + \sum_{j=1}^{k} \sum_{i=1}^{n_j} \log h(\|\mathbf{T}_{ij}\|^2), & \text{no caso (C)}, \end{cases}$$

onde $\mathbf{T}$, $\mathbf{T}_{(j)}$ e $\mathbf{T}_{ij}$ são como em (4.44). Note que

$$\|\mathbf{T}\|^2 = \sum_{j=1}^{k} \|\mathbf{T}_{(j)}\|^2 = \sum_{j=1}^{k} \sum_{i=1}^{n_j} \|\mathbf{T}_{ij}\|^2. \tag{4.56}$$

Assim, se $\boldsymbol{\theta}_j = (\theta_{1j}, \theta_{2j}, \theta_{3j}, \theta_{4j}, \beta_j)^\top$, $j = 1, \ldots, k$ (como em 4.15), temos que

$$\frac{\partial L}{\partial \theta_{lj}} = \begin{cases} -\frac{n_j}{2} \mathrm{tr}\left(\boldsymbol{\Sigma}_j^{-1} \frac{\partial \boldsymbol{\Sigma}_j}{\partial \theta_{lj}}\right) + W_f(\|\mathbf{T}\|^2) \frac{\partial}{\partial \theta_{lj}} \|\mathbf{T}\|^2, & \text{no caso (A)}, \\ -\frac{n_j}{2} \mathrm{tr}\left(\boldsymbol{\Sigma}_j^{-1} \frac{\partial \boldsymbol{\Sigma}_j}{\partial \theta_{lj}}\right) + W_g(\|\mathbf{T}_{(j)}\|^2) \frac{\partial}{\partial \theta_{lj}} \|\mathbf{T}_{(j)}\|^2, & \text{no caso (B)}, \\ -\frac{n_j}{2} \mathrm{tr}\left(\boldsymbol{\Sigma}_j^{-1} \frac{\partial \boldsymbol{\Sigma}_j}{\partial \theta_{lj}}\right) + \sum_{i=1}^{n_j} W_h(\|\mathbf{T}_{ij}\|^2) \frac{\partial}{\partial \theta_{lj}} \|\mathbf{T}_{ij}\|^2, & \text{no caso (C)}, \end{cases}$$

$l = 1, 2, 3, 4, 5$, onde

$$W_k(u) = \frac{\mathrm{d}}{\mathrm{d}u} \log k(u) = \frac{k'(u)}{k(u)}, \ u \geq 0, \ k = f, g, h,$$

$$\frac{\partial \|\mathbf{T}\|^2}{\partial \theta_{lj}} = -2 \frac{\partial \boldsymbol{\mu}^\top}{\partial \theta_{lj}} \boldsymbol{\Sigma}^{-1/2} \mathbf{T} - \mathbf{T}^\top \boldsymbol{\Sigma}^{-1/2} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_{lj}} \boldsymbol{\Sigma}^{-1/2} \mathbf{T},$$

$$\frac{\partial \|\mathbf{T}_{(j)}\|^2}{\partial \theta_{lj}} = -2 \frac{\partial \boldsymbol{\mu}_{(j)}^\top}{\partial \theta_{lj}} \boldsymbol{\Sigma}_{(j)}^{-1/2} \mathbf{T}_{(j)} - \mathbf{T}_{(j)}^\top \boldsymbol{\Sigma}_{(j)}^{-1/2} \frac{\partial \boldsymbol{\Sigma}_{(j)}}{\partial \theta_{lj}} \boldsymbol{\Sigma}_{(j)}^{-1/2} \mathbf{T}_{(j)}$$

e

$$\frac{\partial \|\mathbf{T}_{ij}\|^2}{\partial \theta_{lj}} = -2 \frac{\partial \boldsymbol{\mu}_j^\top}{\partial \theta_{lj}} \boldsymbol{\Sigma}_j^{-1/2} \mathbf{T}_{ij} - \mathbf{T}_{ij}^\top \boldsymbol{\Sigma}_j^{-1/2} \frac{\partial \boldsymbol{\Sigma}_j}{\partial \theta_{lj}} \boldsymbol{\Sigma}_j^{-1/2} \mathbf{T}_{ij}.$$

De (4.56) segue que

$$\frac{\partial \|\mathbf{T}\|^2}{\partial \theta_{lj}} = \frac{\partial \|\mathbf{T}_{(j)}\|^2}{\partial \theta_{lj}} = \sum_{i=1}^{n_j} \frac{\partial \|\mathbf{T}_{ij}\|^2}{\partial \theta_{lj}}. \tag{4.57}$$

Também, de (4.42), temos que

$$\frac{\partial \boldsymbol{\Sigma}_{(j)}}{\partial \theta_{lj}} = \mathbf{I}_{n_j} \otimes \frac{\partial \boldsymbol{\Sigma}_j}{\partial \theta_{lj}}, \quad \frac{\partial \boldsymbol{\Sigma}_{(*)}}{\partial \theta_{lj}} = \operatorname{Diag}\left(\mathbf{0}, \ldots, \mathbf{I}_{n_j} \otimes \frac{\partial \boldsymbol{\Sigma}_j}{\partial \theta_{lj}}, \ldots, \mathbf{0}\right),$$

$$\frac{\partial \boldsymbol{\mu}_{(j)}}{\partial \theta_{lj}} = \mathbf{1}_{n_j} \otimes \frac{\partial \mu_j}{\partial \theta_{lj}}, \quad \frac{\partial \boldsymbol{\mu}_{(*)}}{\partial \theta_{lj}} = \operatorname{Diag}\left(\mathbf{0}^{\top}, \ldots, \mathbf{1}_{n_j}^{\top} \otimes \frac{\partial \boldsymbol{\mu}_j^{\top}}{\partial \boldsymbol{\theta}_{lj}}, \ldots, \mathbf{0}^{\top}\right)$$

e, conseqüentemente, de (4.42) e (4.43), segue que

i) $\boldsymbol{\Sigma}_{(*)}^{-1} \dfrac{\partial \boldsymbol{\Sigma}_{(*)}}{\partial \theta_{lj}} = \operatorname{Diag}\left(\mathbf{0}, \ldots, \boldsymbol{\Sigma}_{(j)}^{-1} \dfrac{\partial \boldsymbol{\Sigma}_{(j)}}{\partial \theta_{lj}}, \ldots, \mathbf{0}\right) = \operatorname{Diag}\left(\mathbf{0}, \ldots, \mathbf{I}_{n_j} \otimes \boldsymbol{\Sigma}_j^{-1} \dfrac{\partial \boldsymbol{\Sigma}_j}{\partial \theta_{lj}}, \ldots, \mathbf{0}\right)$

ii) $\dfrac{\partial \boldsymbol{\mu}_{(*)}^{\top}}{\partial \theta_{lj}} \boldsymbol{\Sigma}_{(*)}^{-1} \dfrac{\partial \boldsymbol{\mu}_{(*)}}{\partial \theta_{mj}} = \dfrac{\partial \boldsymbol{\mu}_{(j)}^{\top}}{\partial \theta_{lj}} \boldsymbol{\Sigma}_{(j)}^{-1} \dfrac{\partial \boldsymbol{\mu}_{(j)}}{\partial \theta_{mj}} = n_j \dfrac{\partial \boldsymbol{\mu}_j^{\top}}{\partial \theta_{lj}} \boldsymbol{\Sigma}_j^{-1} \dfrac{\partial \boldsymbol{\mu}_j}{\partial \theta_{mj}}$

iii) $\operatorname{tr}\left(\boldsymbol{\Sigma}_{(*)}^{-1} \dfrac{\partial \boldsymbol{\Sigma}_{(*)}}{\partial \theta_{lj}}\right) = \operatorname{tr}\left(\boldsymbol{\Sigma}_{(j)}^{-1} \dfrac{\partial \boldsymbol{\Sigma}_{(j)}}{\partial \theta_{lj}}\right) = n_j \operatorname{tr}\left(\boldsymbol{\Sigma}_j^{-1} \dfrac{\partial \boldsymbol{\Sigma}_j}{\partial \theta_{lj}}\right)$

iv) $\operatorname{tr}\left(\boldsymbol{\Sigma}_{(*)}^{-1} \dfrac{\partial \boldsymbol{\Sigma}_{(*)}}{\partial \theta_{lj}} \boldsymbol{\Sigma}_{(*)}^{-1} \dfrac{\partial \boldsymbol{\Sigma}_{(*)}}{\partial \theta_{mj}}\right) = \operatorname{tr}\left(\boldsymbol{\Sigma}_{(j)}^{-1} \dfrac{\partial \boldsymbol{\Sigma}_{(j)}}{\partial \theta_{lj}} \boldsymbol{\Sigma}_{(j)}^{-1} \dfrac{\partial \boldsymbol{\Sigma}_{(j)}}{\partial \theta_{mj}}\right) = n_j \operatorname{tr}\left(\boldsymbol{\Sigma}_j^{-1} \dfrac{\partial \boldsymbol{\Sigma}_j}{\partial \theta_{lj}} \boldsymbol{\Sigma}_j^{-1} \dfrac{\partial \boldsymbol{\Sigma}_j}{\partial \theta_{mj}}\right)$

Para $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^{\top}, \ldots, \boldsymbol{\theta}_k^{\top})^{\top}$, seja

$$K(\boldsymbol{\theta}) = \left(E\left[\frac{\partial L}{\partial \boldsymbol{\theta}_j} \frac{\partial L}{\partial \boldsymbol{\theta}_l^{\top}}\right]\right), \quad j, l = 1, \ldots, k, \tag{4.58}$$

a matriz de informação do modelo estrutural elíptico definido segundo as especificaçẽos (A), (B) ou (C). Então, usando a simetria das distribuições de $\mathbf{T}$, $\mathbf{T}_{(j)}$ e $\mathbf{T}_{ij}$ juntamente com (A.13) e (A.14), do Apêndice A, segue após algumas manipulações algébricas que

$$\mathbf{K}(\boldsymbol{\theta}) = \operatorname{Diag}[\mathbf{K}^{(1)}, \ldots, \mathbf{K}^{(k)}], \tag{4.59}$$

onde para cada $j = 1, \ldots, k$,

$$\mathbf{K}^{(j)} = \mathbf{K}(\boldsymbol{\theta}_j) = (K_{l,m}^{(j)}) = \left(E\left[\frac{\partial L}{\partial \theta_{lj}} \frac{\partial L}{\partial \theta_{mj}}\right]\right), \quad l, m = 1, \ldots, 5, \tag{4.60}$$

com

$$\begin{aligned} K_{l,m}^{(j)} &= \frac{4n_j}{p} a_k(2,1) \frac{\partial \boldsymbol{\mu}_j^{\top}}{\partial \theta_{lj}} \boldsymbol{\Sigma}_j^{-1} \frac{\partial \boldsymbol{\mu}_j}{\partial \theta_{mj}} \\ &\quad + \frac{n_j p}{2} \left\{\frac{1}{p(p+1)} a_k(2,2) - \frac{1}{4}\right\} \operatorname{tr}\left(\boldsymbol{\Sigma}_j^{-1} \frac{\partial \boldsymbol{\Sigma}_j}{\partial \theta_{lj}}\right) \operatorname{tr}\left(\boldsymbol{\Sigma}_j^{-1} \frac{\partial \boldsymbol{\Sigma}_j}{\partial \theta_{mj}}\right) \\ &\quad + \frac{2n_j}{p(p+2)} a_k(2,2) \operatorname{tr}\left(\boldsymbol{\Sigma}_j^{-1} \frac{\partial \boldsymbol{\Sigma}_j}{\partial \theta_{lj}} \boldsymbol{\Sigma}_j^{-1} \frac{\partial \boldsymbol{\Sigma}_j}{\partial \theta_{mj}}\right), \end{aligned}$$

$$a_k(r,s) = \begin{cases} E[(W_f(\|\mathbf{T}\|^2))^r \|\mathbf{T}\|^{2s}], & \text{no caso (A)}, \\ E[(W_g(\|\mathbf{T}_{(j)}\|^2))^r \|\mathbf{T}_{(j)}\|^{2s}], & \text{no caso (B)}, \\ E[(W_h(\|\mathbf{T}_{ij}\|^2))^r \|\mathbf{T}_{ij}\|^{2s}], & \text{no caso (C)}, \end{cases} \tag{4.61}$$

$r, s = 1, 2,\ r \geq s$, e

$$p = \begin{cases} 2n, & \text{no caso (A)}, \\ 2n_j, & \text{no caso (B)}, \\ 2, & \text{no caso (C)}. \end{cases} \tag{4.62}$$

Para facilitar a inferência nos modelos elípticos, vamos considerar a parametrização ortogonal que foi derivada para o modelo normal (seção 2), isto é, para cada $j = 1, \ldots, k$,

$$\boldsymbol{\phi}_j = \boldsymbol{\phi}(\boldsymbol{\theta}_j) = (\phi_{1j}, \phi_{2j}, \phi_{3j}, \phi_{4j}, \beta_j)^\top,$$

onde $\phi_{lj}$, $l = 1, \ldots, 5$, é como em (4.20). Daí, segue que

$$\boldsymbol{\mu}_j = \boldsymbol{\mu}(\boldsymbol{\phi}_{L_j}) \quad \text{e} \quad \boldsymbol{\Sigma}_j = \boldsymbol{\Sigma}(\boldsymbol{\phi}_{E_j}), \tag{4.63}$$

o que implica que

$$\boldsymbol{\phi}_{L_j} = (\phi_{1_j}, \phi_{2j})^\top \quad \text{e} \quad \boldsymbol{\phi}_{E_j} = (\phi_{3j}, \phi_{4j}, \beta_j)^\top$$

são ortogonais, isto é, a ortogonalidade de $\boldsymbol{\phi}_{E_j}$ sob o modelo normal é preservada nos modelos elípticos sob qualquer uma das três especificações. Agora, desde que $|\boldsymbol{\Sigma}_j| = \phi_{3j}\phi_{4j}$, temos que

$$\operatorname{tr}\left(\boldsymbol{\Sigma}_j \frac{\partial \boldsymbol{\Sigma}_j}{\partial \beta_j}\right) = \operatorname{tr}\left(\boldsymbol{\Sigma}_j^{-1} \frac{\partial \boldsymbol{\Sigma}_j}{\partial \beta_j} \boldsymbol{\Sigma}_j^{-1} \frac{\partial \boldsymbol{\Sigma}_j}{\partial \phi_{lj}}\right) = 0,\ l = 3, 4, \tag{4.64}$$

o que implica que, para $\beta_j$ é ortogonal a $(\phi_{3j}, \phi_{4j})^\top$, $j = 1, \ldots, k$. Contudo, nos modelos elípticos a ortogonalidade entre $\phi_{3j}$ e $\phi_{4j}$ não é necessariamente válida (veja capítulo 2, seção 2.5).

Seja $\mathbf{K} = \mathbf{K}(\boldsymbol{\phi})$, $\boldsymbol{\phi} = (\boldsymbol{\phi}_1^\top, \ldots, \boldsymbol{\phi}_k^\top)^\top$ a matriz de informação sob a parametrização ortogonal, de modo que de (4.60), (4.63) e (4.64), temos que

$$\mathbf{K} = \operatorname{Diag}(\mathbf{K}^{(1)}, \ldots, \mathbf{K}^{(k)}), \tag{4.65}$$

onde agora $\mathbf{K}^{(j)} = \mathbf{K}(\boldsymbol{\phi}_j) = \mathrm{Diag}(\mathbf{K}_{L_j}, \mathbf{K}_{E_j})$ é a matriz de informação correspondente à $j$-ésima população, com $\mathbf{K}_{L_j}$ e $\mathbf{K}_{E_j}$ sendo as submatrizes de informação relativas aos parâmetros $\boldsymbol{\phi}_{L_j}$ e $\boldsymbol{\phi}_{E_j}$, respectivamente. Estas matrizes são dadas por

$$K_{L_j} = \frac{4}{p} a_k(2,1) \left( \frac{1}{n_j} \boldsymbol{\Sigma}_j \right)^{-1}, \quad j = 1, \ldots, k, \tag{4.66}$$

e

$$\mathbf{K}_{E_j} = (K^{(j)}_{l,m}), \quad l, m = 3, 4, 5, \tag{4.67}$$

onde

$$K^{(j)}_{l,m} = \begin{cases} \frac{p}{2} \left[ \frac{p+4\delta_{lm}}{p^2(p+2)} a_k(2,2) - \frac{1}{4} \right] \left( \frac{\phi_{lj}\phi_{mj}}{n_j} \right)^{-1}, & l, m = 3, 4, \\ \frac{4}{p(p+2)} a_k(2,2) \left( \frac{\sigma^2_{\beta_j}}{n_j} \right)^{-1}, & l, m = 5 \ (\phi_{5_j} = \beta_j), \\ 0, & l = 3, 4, \ m = 5 \text{ ou } l = 5, \ m = 3, 4, \end{cases}$$

com $\sigma^2_{\beta_j}$ como em (4.25), $\delta_{lm}$ é o delta de Kronecker, $a_k(2,2)$ e $p$ são como em (4.61) e (4.62), respectivamente.

## 4.4 Testes assintóticos modificados por correções tipo Bartlett

### 4.4.1 Introdução

A utilidade dos testes assintóticos, como por exemplo as estatísticas da razão de verossimilhança ($G$), escore ($S$) e de Wald ($W$), podem ser limitadas em algumas situações práticas, uma vez que elas requerem amostras suficientemente grandes para que a distribuição da estatística do teste seja mais próxima da distribuição da estatística de referência. Contudo, em problemas que envolvem uma função de verossimilhança regular, isto é, quando é permitida a troca de derivadas da função de verossimilhança com respeito aos parâmetros com integrais sobre o espaço amostral, a aproximação das distribuições das estatísticas $G$ e $S$ com a distribuição de referência qui-quadrado pode ser melhorada em modelos com distribuição amostral contínua, multiplicando-se estas estatísticas por fatores de correção do tipo Bartlett. Este procedimento foi originalmente proposto por Bartlett (1937) para o teste de homogeneidade de variâncias em populações normais. Posteriormente, Lawley (1956) desenvolve o fator de Bartlett da estatística $G$ num contexto bem geral. Outras referências neste contexto são, por exemplo, Hayakawa (1977), Barndorff, Nielsen e Cox (1984), Cordeiro (1987) e McCullagh e Cox (1986). Recentemente, Cordeiro e Ferrari (1991) mostram que o procedimento também pode ser aplicado à estatística $S$, mas neste caso o fator de correção obtido é uma função polinomial de segundo grau na própria estatística. A conclusão comum destes autores é que para amostras finitas, em particular em amostras de tamanho moderado, as estatísticas corrigidas deverão ter um melhor desempenho que as respectivas estatísticas originais, uma vez que as estatísticas modificadas têm distribuições mais próximas da distribuição qui-quadrado de referência que a não corrigida.

Para as estatísticas corrigidas, tal aproximação é garantida pelo menos até ordem $n^{-1}$ ($n$ tamanho da amostra), no sentido de que sob a hipótese nula $H_0$ as funções de distribuição acumulada (FDA) destas estatísticas coincidem com FDA qui-quadrado de referência, até ordem $n^{-1}$ (ignorando os termos de ordem inferior a $n^{-1}$).

Os fatores de correção destas estatísticas são funções de cumulantes conjuntos de derivadas (com respeito aos parâmetros) até quarta ordem do logaritmo da função de verossimilhança, de modo que suas derivações podem envolver manipulações algébricas muito trabalhosas em alguns problemas particulares. Este fato tem motivado diferentes trabalhos com o propósito de obter fórmulas para estes fatores de correção que sejam de fácil implementação computacional e que englobe um número amplo de modelos probabilísticos.

A aproximação da distribuição da estatística da razão de máxima verossimilhança $G$ sob a hipótese nula $H_0$ para a distribuição qui-quadrado pode ser melhorada utilizando o fator de Bartlett via o procedimento de Lawley (1956), que muitas vezes é um procedimento que requer um processo extenso de manipulações algébricas. A principal contribuição desta seção é mostrar que este fator sob o modelo normal pode ser obtido para algumas hipóteses avaliando-se diretamente o valor esperado de $G$ sob a hipótese nula $H_0$, onde a esperança de $G$ sob a hipótese nula será denotado por $E_0[G]$. Extensões são consideradas para os modelos elípticos segundo os enfoques (A) e (B). Apesar deste método nem sempre ser aplicável, pois depende de uma variável aleatória $T$, que nem sempre tem uma distribuição fechada que permita a determinação de $E_0[\log T]$. Contudo, o fator de Bartlett, usando esta nova metodologia considerada em Arellano-Valle e Bolfarine (1995), não depende dos parâmetros do modelo em situações onde pode ser aplicado.

A seguir apresentamos, de forma resumida, o procedimento de Lawley (1956) para a obtenção do fator de correção de Bartlett. Seja $\mathbf{Y}$ um vetor aleatório de dimensão $n$, com logaritmo da função de verossimilhança $L(\boldsymbol{\theta}) = \log P_Y(\mathbf{Y}|\boldsymbol{\theta})$, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^\top$ e seja $\widehat{\boldsymbol{\theta}}$ o EMV de $\boldsymbol{\theta}$. Então, Lawley (1956) expande a função $2\{L(\widehat{\boldsymbol{\theta}}) - L(\boldsymbol{\theta})\}$ em torno de $\boldsymbol{\theta}$ e mostra que

$$E[2\{L(\widehat{\boldsymbol{\theta}}) - L(\boldsymbol{\theta})\}] = p + e_p + O(n^{-2}), \tag{4.68}$$

onde $e_p$ é de ordem $n^{-1}$ e depende de cumulantes de derivadas do logaritmo da função de verossimilhança referentes ao vetor $\boldsymbol{\theta}$.

Considerando a partição $\boldsymbol{\theta}$ como $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$, onde $\boldsymbol{\theta}_2$ é um vetor de dimensão $q$, o interesse é em desenvolver a correção da razão de verossimilhança para testar a hipótese composta $H_0 : \boldsymbol{\theta}_2 = \boldsymbol{\theta}_0$ contra $H_1 : \boldsymbol{\theta}_2 \neq \boldsymbol{\theta}_0$. Seja $\widetilde{\boldsymbol{\theta}}$ o EMV sob $H_0$. Então, procedendo similarmente para $2\{L(\widetilde{\boldsymbol{\theta}}) - L(\boldsymbol{\theta})\}$, temos que

$$E[2\{L(\widetilde{\boldsymbol{\theta}}) - L(\boldsymbol{\theta})\}] = p - q + e_{p-q} + O(n^{-2}), \tag{4.69}$$

onde $e_{p-q}$ é de ordem $n^{-1}$ e depende dos cumulantes referentes ao vetor $\boldsymbol{\theta}_1$ de dimensão $p - q$. Se $G$ é a estatística da razão de verossimilhança para o teste de $H_0 : \boldsymbol{\theta}_2 = \boldsymbol{\theta}_0$ contra

$H_1 : \boldsymbol{\theta}_2 \neq \boldsymbol{\theta}_0$, então, notando que

$$G = 2\{L(\widehat{\boldsymbol{\theta}}) - L(\boldsymbol{\theta})\} - 2\{L(\widetilde{\boldsymbol{\theta}}) - L(\boldsymbol{\theta})\} \tag{4.70}$$

e tomando a esperança sob a hipótese nula $H_0$, segue de (4.68), (4.69) e (4.70) que

$$E_0[G] = q(1+d) + O(n^{-2}), \tag{4.71}$$

onde

$$d = \frac{1}{q}(e_p - e_{p-q}) \tag{4.72}$$

é um termo de ordem $n^{-1}$ ($d = O(n^{-1})$). Assim, segue que a estatística $G$ vezes o fator de Bartlett $(1+d)^{-1}$ ou $(1-d)$, isto é,

$$G^* = (1+d)^{-1}G \quad \text{ou} \quad G^* = (1-d)G \tag{4.73}$$

é tal que sob a hipótese nula, $E_0[G^*] = q + O(n^{-2})$. Note que sob $H_0$, $E_0[G] = q + O(n^{-1})$ e $G^*$ tem distribuição assintótica $\chi^2_q$, uma vez que $G$ tem distribuição assintótica $\chi^2_q$. Lawley (1956) também mostra que sob $H_0$, $E_0[G^{*m}] = \mu_m + O(n^{-2})$, $m = 1, 2, \ldots$ onde $\mu_m = q(q+2)\cdots(q+2(m-1))$ é o momento de ordem $m$ da distribuição $\chi^2_q$, enquanto que $E[G^m] = \mu_m + O(n^{-1})$, $m = 1, 2, \ldots$ Isto prova que sob $H_0$ a estatística $G^*$ tem seus momentos mais próximos dos respectivos momentos da estatística de referência $\chi^2_q$ que a estatística não corrigida $G$.

A obtenção do valor esperado de $G$ sob a hipótese nula $H_0$ é baseada na função digama, que é definida a seguir: Seja $V$ uma variável aleatória com $V \sim Ga(a, b)$, onde $Ga(a, b)$ é função gama com parâmetros $a$ e $b$, tal que $E[V] = a/b$ e $\text{Var}[V] = a/b^2$. Então $E[\log(V)] = \psi(a) - \log b$, com $\psi$ sendo a função digama (veja Abramowitz e Stegun, 1965). Algumas de suas propriedades são:

$$\begin{aligned} \psi(m) &= \psi(m-1) + \frac{1}{m-1}, & (4.74)\\ \psi(2m) &= \frac{1}{2}\psi(m) + \frac{1}{2}\psi\left(m + \frac{1}{2}\right) + \log 2, & (4.75)\\ \psi(m) &= \log m - \frac{1}{2m} - \frac{1}{12m^2} + O(m^{-4}). & (4.76) \end{aligned}$$

**Lema 4.1** *Seja $\psi$ a função digama e $n = n_1 + \cdots + n_k$. Então*

*i)* $\psi\left(\frac{n_j - 1}{2}\right) - \log\frac{n_j}{2} = \frac{1}{n_j}\left[-2 - \frac{5}{6n_j} + O(n_j^{-2}\right]$, $j = 1, \ldots, k$,

*ii)* $\psi\left(\frac{n-k-1}{2}\right) + \psi\left(\frac{n-k}{2}\right) - 2\log\frac{n}{2} = \frac{1}{n}\left[-2k - 3 - \frac{6(k+1)(k+2)+1}{6n}\right.$
$\left.+O(n^{-2})\right]$.

**Prova:** i) De (4.74) e (4.76), temos que

$$\psi\left(\frac{n_j - 1}{2}\right) - \log\frac{n_j}{2} = \log\left(\frac{n_j + 1}{n_j}\right) - \frac{1}{n_j + 1} - \frac{2}{n_j - 1} - \frac{1}{3(n_j + 1)^2} + O(n_j^{-4})$$
$$= \log\left(1 + \frac{1}{n_j}\right) - \frac{1}{n_j}\left[\left(1 + \frac{1}{n_j}\right)^{-1} + 2\left(1 - \frac{1}{n_j}\right)^{-1}\right] - \frac{1}{3n_j^2}\left(1 + \frac{1}{n_j}\right)^{-1} + O(n_j^{-4}).$$

Agora, das relações (veja Arellano-Valle, 1994),

$$\begin{aligned} \log\left(1 + \frac{1}{m}\right) &= \frac{1}{m} + \frac{1}{2m^2} + O(m^{-3}), \\ \left(1 + \frac{1}{m}\right)^{-r} &= 1 - \frac{r}{m} + O(m^{-2}), \quad r = 1, 2, \\ \text{e} \qquad \left(1 - \frac{a}{m}\right)^{-1} &= 1 + \frac{a}{m} + O(m^{-2}), \end{aligned}$$

temos que

$$\begin{aligned} \psi\left(\frac{n_j - 1}{2}\right) - \log\frac{n_j}{2} &= \frac{1}{n_j} + \frac{1}{2n_j^2} + O(n_j^{-3}) - \frac{1}{n_j}\left[3 + \frac{1}{n_j} + O(n_j^{-2}\right] \\ &\quad -\frac{1}{3n_j^2}\left[1 - \frac{1}{n_j} + O(n_j^{-2})\right] + O(n_j^{-4}) \\ &= \frac{1}{n_j}\left[-2 - \frac{5}{6n_j} + O(n_j^{-2})\right]. \end{aligned}$$

ii) De (4.74), (4.75) e (4.76), segue, após algumas manipulações algébricas, que

$$\psi\left(\frac{n-k-1}{2}\right) + \psi\left(\frac{n-k}{2}\right) - 2\log\frac{n}{2} =$$
$$= \psi\left(\frac{n-k}{2} + \frac{1}{2}\right) + \psi\left(\frac{n-k}{2}\right) - \frac{2}{n-k-1} - 2\log\frac{n}{2}$$

$$
\begin{aligned}
&= 2\psi(n-k) - 2\log 2 - \frac{2}{n-k-1} - 2\log\frac{n}{2} \\
&= 2\left[\psi(n) - \sum_{j=1}^{k}\frac{1}{n-j}\right] - \frac{2}{n-k-1} - 2\log n \\
&= 2\psi(n) - 2\sum_{j=1}^{k+1}\frac{1}{n-j} - 2\log n \\
&= 2\log n - \frac{1}{n} - \frac{1}{6n^2} - 2\sum_{j=1}^{k+1}\frac{1}{n-j} - 2\log n + O(n^{-4}) \\
&= \frac{1}{n}\left[-1 - \frac{1}{6n} - 2\sum_{j=1}^{k+1}\left(1-\frac{j}{n}\right)^{-1} + O(n^{-3})\right] \\
&= \frac{1}{n}\left[-1 - \frac{1}{6n} - 2\sum_{j=1}^{k+1}\left(1+\frac{j}{n} + O(n^{-2})\right) + O(n^{-3})\right] \\
&= \frac{1}{n}\left[-1 - 2(k+1) - \frac{(k+1)(k+2)}{n} - \frac{1}{6n} + O(n^{-2})\right] \\
&= \frac{1}{n}\left[-2k - 3 - \frac{6(k+1)(k+2)+1}{6n} + O(n^{-2})\right].
\end{aligned}
$$

Notemos que se $n = n_1 + \cdots + n_k$, então

$$
\sum_{j=1}^{k} O(n_j^{-2}) + O(n^{-2}) = O(n_*^{-2}), \tag{4.77}
$$

onde $n_* = \min_{1\le j\le k}\{n_j\}$. Com efeito,

$$
\begin{aligned}
\sum_{j=1}^{k} O(n_j^{-2}) + O(n^{-2}) &= O\left(\max_{1\le j\le k}\{n_j^{-2}\}\right) + O(n^{-2}) \\
&= O\left(\left(\min_{1\le j\le k}\{n_j\}\right)^{-2}\right) + O(n^{-2}) \\
&= O\left(\max\{n_*^{-2}, n^{-2}\}\right) \\
&= O\left((\min\{n_*, n\})^{-2}\right) \\
&= O(n_*^{-2}).
\end{aligned}
$$

### 4.4.2 O teste da razão de verossimilhança para a hipótese $H_0$ : $\beta_j = \beta_{0j}$, $j = 1, \dots, k$

Consideremos primeiramente o modelo definido em (4.1) sob normalidade e sob a parametrização ortogonal definida em (4.20). Suponha que deseja-se testar a hipótese nula $H_0 : \beta_j = \beta_{j0}$, onde $\beta_{j0}$ conhecido, $j = 1, \dots, k$, usando a estatística da razão de verossimilhança

$$G = 2(L(\widehat{\boldsymbol{\phi}}) - L(\widetilde{\boldsymbol{\phi}})), \tag{4.78}$$

onde $\widehat{\boldsymbol{\phi}}$ e $\widetilde{\boldsymbol{\phi}}$ são os EMV sob o modelo não restrito e restrito (sob $H_0$) de $\boldsymbol{\phi}$, respectivamente, onde $\boldsymbol{\phi} = (\boldsymbol{\phi}_1^\top, \dots, \boldsymbol{\phi}_k^\top)^\top$, com $\boldsymbol{\phi}_j = (\phi_{1j}, \phi_{2j}, \phi_{3j}, \phi_{4j}, \beta_j)^\top$ como em (4.20) que pode ser escrito por $\boldsymbol{\phi}_j = (\boldsymbol{\phi}_{L_j}^\top, \boldsymbol{\phi}_{E_j}^\top)^\top$, onde $\boldsymbol{\phi}_{L_j}$ e $\boldsymbol{\phi}_{E_j}$ são como em (4.21). Então, sob o modelo normal segue que no modelo não restrito, os EMV $\widehat{\phi}_{lj}$ de $\phi_{lj}$, $l = 1, \dots, 5$, são dados por (4.28), (4.29) e (4.30). Por outro lado, os EMV sob $H_0$ de $\phi_{lj}$, $l = 1, \dots, 4$, são dados por

$$\widetilde{\boldsymbol{\phi}}_{L_j} = \widehat{\boldsymbol{\phi}}_{L_j} = \overline{\mathbf{Z}}_j \quad \text{e} \quad \widetilde{\phi}_{lj} = \overline{\boldsymbol{\alpha}}_{lj}^\top(\beta_{0j})\mathbf{S}_{n_j}\overline{\boldsymbol{\alpha}}_{lj}(\beta_{0j}), \; l = 3, 4, \tag{4.79}$$

Agora, de (4.32), temos que sob $H_0$, $\mathbf{S}_{n_j} \sim W_2\left(n_j - 1, \frac{1}{n_j}\boldsymbol{\Sigma}(\boldsymbol{\phi}_{E_j}^0)\right)$, com $\boldsymbol{\phi}_{E_j}^0 = (\phi_{3j}, \phi_{4j}, \beta_{0j})^\top$; então, das propriedades da matriz Wishart(Muirhead, 1982) segue que

$$n_j \frac{\widetilde{\phi}_{lj}}{\phi_{lj}} \sim \chi^2_{n_j - 1}, \quad l = 3, 4. \tag{4.80}$$

Por outro lado, temos que sob a hipótese nula $H_0$ a estatística da razão de verossimilhança $G$ (definida em (4.78)) tem distribuição assintótica $\chi^2_k$.

**Teorema 4.1** *Sob o modelo normal e a condição de identificabilidade $\lambda_{e_j}$ ou $\lambda_{x_j}$, $j = 1, \dots, k$, conhecidos, a estatística da razão de verossimilhança $G$ e sua estatística corrigida $G^*$ para testar $H_0$ são dadas por*

$$G = \sum_{j=1}^{k} n_j \left(\log(\widetilde{\phi}_{3j}\widetilde{\phi}_{4j}) - \log(\widehat{\phi}_{3j}\widehat{\phi}_{4j})\right) \quad e \quad G^* = (1 + d)^{-1}G, \tag{4.81}$$

*respectivamente, onde*

$$d = \sum_{j=1}^{k} \frac{5}{2n_j k}. \tag{4.82}$$

**Prova:** Substituindo os estimadores $\widehat{\phi}_j$ e $\widetilde{\phi}_j$ dados em (4.78) e após algumas manipulações algébricas, temos

$$\begin{aligned} G &= \sum_{i=1}^{k} n_j \left[\log|\Sigma(\widetilde{\phi}_{E_j})| - \log|\Sigma(\widehat{\phi}_{E_j})|\right] \\ &= \sum_{j=1}^{k} G_j, \end{aligned}$$

onde

$$G_j = n_j \left[\log(\widetilde{\phi}_{3j}\widetilde{\phi}_{4j}) - \log(\widehat{\phi}_{3j}\widehat{\phi}_{4j})\right]$$

é a estatística da razão de verossimilhança correspondente à $j$-ésima população. Arellano-Valle e Bolfarine (1996) mostram que sob $H_0$

$$E_0[G_j] = n_j \left[\psi\left(\frac{n_j - 1}{2}\right) - \psi\left(\frac{n_j - 2}{2}\right)\right] = 1 + \frac{5n_j}{2} + O(n_j^{-2}) \tag{4.83}$$

cuja demonstração é similar à prova do Lema 4.1. Assim, temos que

$$E_0[G] = \sum_{j=1}^{k} E_0[G_j] = k + \sum_{j=1}^{k} \frac{5}{2n_j} + O(n_*^{-2}),$$

onde $n_* = \min_{1 \leq j \leq k}\{n_j\}$. Então, de (4.73) a estatística corrigida de $G$ é dada por $G^* = (1+d)^{-1}G$, com $d$ como em (4.82). Assim, $G^*$ tem distribuição mais próxima da distribuição qui-quadrado com $k$ graus de liberdade que a distribuição de $G$. □

Note que o resultado anterior coincide com o resultado de Wong (1991), obtido usando um método diferente de aproximação sob a condição de identificabilidade $\lambda_{e_j}$, $j = 1, \ldots, k$, conhecidos ($\lambda_{e_j} = 1$) e por Bolfarine e Nascimento (1992) sob a condição de identificabilidade $\lambda_{x_j}$, $j = 1, \ldots, k$, conhecidos. Em ambos os resultados o fator de correção é obtido utilizando o procedimento de Lawley (1956) descrito na seção anterior.

Seja $G_E$ a estatística da razão de verossimilhança correspondente ao modelo estrutural elíptico seguindo o enfoque (A), (B) ou (C). Por exemplo, sob o enfoque (A) o logaritmo da função de verossimilhança pode ser escrita como (veja (4.12) para o caso normal)

$$L(\phi) = \sum_{j=1}^{k} \left(-\frac{n_j}{2}\right) \log |\Sigma_j| + \log f\left(\sum_{j=1}^{k} n_j \left\{ \operatorname{tr}\left(\Sigma_j^{-1} \mathbf{S}_{n_j}\right) + (\overline{\mathbf{Z}}_j - \boldsymbol{\mu}_j)^\top \Sigma_j^{-1} (\overline{\mathbf{Z}}_j - \boldsymbol{\mu}_j) \right\}\right). \tag{4.84}$$

Agora, supondo que $f$ é uma função contínua e decrescente sobre $(0, +\infty)$ temos que os EMV de $\boldsymbol{\mu}_j$ e $\Sigma_j$ sob o modelo não restrito são dados por (4.53). Substituindo estes estimadores em (4.84), segue que

$$L(\widehat{\phi}) = \sum_{j=1}^{k} \left(-\frac{n_j}{2}\right) \log \left| \frac{np}{u_f} \widehat{\Sigma}_{N_j} \right| + \log f(u_f). \tag{4.85}$$

Por outro lado, sob a hipótese nula $H_0$, os EMV de $\boldsymbol{\mu}_j$ e $\Sigma_j$ são dados por

$$\widetilde{\boldsymbol{\mu}}_j = \widehat{\boldsymbol{\mu}}_j = \overline{\mathbf{Z}}_j \quad \text{e} \quad \widetilde{\Sigma}_j = \frac{np}{u_f} \widetilde{\Sigma}_{N_j} \tag{4.86}$$

e substituindo-os em (4.84) obtemos

$$L(\widetilde{\phi}) = \sum_{j=1}^{k} \left(-\frac{n_j}{2}\right) \log \left| \frac{np}{u_f} \widetilde{\Sigma}_{N_j} \right| + \log f(u_f). \tag{4.87}$$

Assim, de (4.85) e (4.87) temos que, sob $H_0$,

$$G_E = 2\{L(\widehat{\phi}) - L(\widetilde{\phi})\} = \sum_{j=1}^{k} n_j \log \left( \frac{\widetilde{\phi}_{3j} \widetilde{\phi}_{4j}}{\widehat{\phi}_{3j} \widehat{\phi}_{4j}} \right), \tag{4.88}$$

onde $\widetilde{\phi}_{3j}$, $\widetilde{\phi}_{4j}$, $\widehat{\phi}_{3j}$ e $\widehat{\phi}_{4j}$ são os EMV sob o modelo normal. □

**Corolário 4.1** *Sob o modelo elíptico correspondente à especificação (A) e a condição de identificabilidade $\lambda_{e_j}$ ou $\lambda_{x_j}$, $j = 1, \ldots, k$, conhecidos, a estatística corrigida de $G_E$ é dada por*

$$G_E^* = (1 + d)^{-1} G_E,$$

*onde $d$ é como no Teorema 4.1.*

**Prova:** A demonstração segue de (4.88) e do Teorema 4.1. □

Agora, sob o modelo elíptico segundo o enfoque (B), o logaritmo da função de verossimilhança é dada por

$$L(\phi) = \sum_{j=1}^{k} \left(-\frac{n_j}{2}\right) \log |\Sigma_j| + \sum_{j=1}^{k} \log g\left(n_j \left\{\operatorname{tr}(\Sigma_j^{-1} S_{n_j}) + (\overline{\mathbf{Z}}_j - \mu_j)^\top \Sigma_j^{-1} (\overline{\mathbf{Z}}_j - \mu_j)\right\}\right). \tag{4.89}$$

Assim, supondo que $g$ é contínua e decrescente sobre $(0, +\infty)$ (lembremos que $g(\cdot)$ depende de $j$, pois é a função densidade esférica em $\mathbb{R}^{2n_j}$), temos que os EMV de $\boldsymbol{\mu}_j$ e $\boldsymbol{\Sigma}_j$ sob a hipótese não restrita são dados em (4.55). Substituindo estes estimadores em (4.89), obtemos que

$$L(\widehat{\phi}) = \sum_{i=1}^{k} \left(-\frac{n_j}{2}\right) \log \left|\frac{n_j p}{u_g} \widehat{\Sigma}_{N_j}\right| + \sum_{j=1}^{k} \log g(u_g). \tag{4.90}$$

Similarmente, sob $H_0$, temos que os EMV de $\boldsymbol{\mu}_j$ e $\boldsymbol{\Sigma}_j$ são dados por

$$\widetilde{\boldsymbol{\mu}}_j = \widehat{\boldsymbol{\mu}}_j \quad \text{e} \quad \widetilde{\boldsymbol{\Sigma}}_j = \frac{n_j p}{u_g} \widetilde{\Sigma}_{N_j}, \tag{4.91}$$

respectivamente. Substituindo-os em (4.89), segue que

$$L(\widetilde{\phi}) = \sum_{j=1}^{k} \left(-\frac{n_j}{2}\right) \log \left|\frac{n_j p}{u_g} \widetilde{\Sigma}_{N_j}\right| + \sum_{j=1}^{k} \log g(u_g). \tag{4.92}$$

Portanto,

$$G_E = \sum_{j=1}^{k} n_j \log\left(\frac{\widetilde{\phi}_{3j} \widetilde{\phi}_{4j}}{\widehat{\phi}_{3j} \widehat{\phi}_{4j}}\right), \tag{4.93}$$

a prova segue de (2.62).

**Corolário 4.2** *Sob o modelo elíptico correspondente à especificação (B) e a condição de identificabilidade $\lambda_{e_j}$ ou $\lambda_{x_j}$, $j = 1, \ldots, k$, conhecidos, a estatística corrigida $G_E$ é dada por*

$$G_E^* = (1 + d)^{-1} G_E,$$

*onde $d$ é como no Teorema 4.1.*

**Prova:** A demonstração é análoga à do Corolário 4.1. □

A seguir, sob a condição de identificabilidade $\lambda_{x_j} = \frac{\sigma^2_{x_j}}{\sigma^2_{u_j}}$, $j = 1, \ldots, k$, conhecidos e $\sigma^2_{u_j} = \sigma^2_{0j}$, $j = 1, \ldots, k$, conhecidos, vamos considerar a hipótese nula $H_0 : \beta_j = \beta_{0j}$, $j = 1, \ldots, k$, com $\beta_{0j}$ conhecido sob o modelo normal. Com efeito, sob a parametrização ortogonal

$$\phi_{1j} = \alpha_j + \beta\mu_{x_j}, \quad \phi_{2j} = \mu_{x_j}$$

$$\phi_{3j} = \lambda_{x_j}\beta_j^2\sigma^2_{0j} + (\lambda_{x_j} + 1)\sigma^2_{e_j}$$

$$\phi_{4j} = \beta_j$$

temos que o logaritmo da função de verossimilhança pode ser expressa por

$$L(\boldsymbol{\phi}) = \sum_{i=1}^{k}\left(-\frac{n_j}{2}\right)\log|\boldsymbol{\Sigma}_j| - \frac{1}{2}\sum_{j=1}^{k} n_j\left\{\operatorname{tr}(\boldsymbol{\Sigma}_j^{-1}\mathbf{S}_{n_j}) + (\overline{\mathbf{Z}}_j - \boldsymbol{\mu}_j)^{\top}\boldsymbol{\Sigma}_j^{-1}(\overline{\mathbf{Z}}_j - \boldsymbol{\mu}_j)\right\}, \tag{4.94}$$

onde $|\boldsymbol{\Sigma}_j| = \phi_{3j}\sigma^2_{0j}$, de modo que os EMV sob $H_1$ são dados por

$$\widehat{\mu}_j = \overline{\mathbf{Z}}_j, \quad \widehat{\phi}_{3j} = (\lambda_{x_j} + 1)S^{(j)}_{YY.X} \quad \text{e} \quad \widehat{\beta}_j = \left(\frac{\lambda_{x_j} + 1}{\lambda_{x_j}}\right)\frac{S^{(j)}_{YX}}{S^{(j)}_{XX}}. \tag{4.95}$$

Por outro lado, os EMV sob a hipótese nula $H_0$ são dados por

$$\widehat{\mu}_j = \widehat{\mu}_j \quad \text{e} \quad \widetilde{\phi}_{3j} = (\lambda_{x_j} + 1)S^{(j)}_{YY} - 2(\lambda_{x_j}\beta_{0j})S^{(j)}_{YX} + (\lambda_{x_j}\beta_{0j})^2\frac{S^{(j)}_{XX}}{\lambda_{x_j} + 1}. \tag{4.96}$$

Logo, substituindo estes estimadores em (4.12), temos que

$$\begin{aligned} L(\widehat{\boldsymbol{\phi}}) &= \sum_{j=1}^{k}\left[\left(-\frac{n_j}{2}\right)\log(\widehat{\phi}_{3j}\sigma^2_{0j}) - \frac{n_j}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}_j^{-1}(\widehat{\boldsymbol{\phi}}_{E_j})\mathbf{S}_{n_j}\right)\right] \\ &= \sum_{j=1}^{k}\left[\left(-\frac{n_j}{2}\right)\log(\widehat{\phi}_{3j}\sigma^2_{0j}) - \frac{n_j}{2}\left(1 + \frac{1}{\sigma^2_{0j}}\frac{S^{(j)}_{XX}}{\lambda_{x_j} + 1}\right)\right] \end{aligned}$$

e

$$L(\widetilde{\boldsymbol{\phi}}) = \sum_{j=1}^{k}\left[\left(-\frac{n_j}{2}\right)\log(\widehat{\phi}_{3j}\sigma^2_{0j}) - \frac{n_j}{2}\left(1 + \frac{1}{\sigma^2_{0j}}\frac{S^{(j)}_{XX}}{\lambda_{x_j} + 1}\right)\right],$$

onde

$$n_j \frac{\widehat{\phi}_{3j}}{\phi_{3j}} \sim \chi^2_{n_j-2} \quad \text{e} \quad n_j \frac{\widetilde{\phi}_{3j}}{\phi_{3j}} \sim \chi^2_{n_j-1}. \tag{4.97}$$

Assim, de (4.78) e das igualdades acima, segue que

$$G = \sum_{j=1}^{k} n_j \log\left(\frac{\widetilde{\phi}_{3j}}{\widehat{\phi}_{3j}}\right). \tag{4.98}$$

Portanto, sob a hipótese nula $H_0$, temos de (4.97) e da demonstração do Teorema 4.1 que

$$\begin{aligned} E_0[G] &= \sum_{j=1}^{k} n_j \left\{ E_0\left[\log\left(\frac{\widetilde{\phi}_{3j}}{\phi_{3j}}\right)\right] - E_0\left[\log\left(\frac{\widehat{\phi}_{3j}}{\phi_{3j}}\right)\right]\right\} \\ &= \sum_{j=1}^{k} n_j \left\{\psi\left(\frac{n_j-1}{2}\right) - \psi\left(\frac{n_j-2}{2}\right)\right\} \\ &= \sum_{j=1}^{k}\left(1 + \frac{5}{2n_j} + O(n_j^{-2})\right) \\ &= k + \sum_{j=1}^{k}\frac{5}{2n_j} + O(n_*^{-2}). \end{aligned}$$

Assim, a estatística corrigida $G^*$ de $G$ é dada como no Teorema 4.1.

O teste acima foi estudado por Wong (1994) sob a condição de identificabilidade $\lambda_{e_j} = \frac{\sigma^2_{e_j}}{\sigma_{u_j}}$, $j = 1, \ldots, k$, conhecidos ($\lambda_{e_j} = 1$) e $\sigma^2_{0j}$, $j = 1, \ldots, k$, conhecidos, usando um método diferente de aproximação.

Nas seções seguintes, examinaremos os testes da razão de verossimilhança para as hipóteses

$$H_0 \; : \; \mathbf{\Sigma}_j = \mathbf{\Sigma}, \quad j = 1, \ldots, k, \tag{4.99}$$

$$H_0 \; : \; \boldsymbol{\mu}_j = \boldsymbol{\mu} \quad \text{e} \quad \mathbf{\Sigma}_j = \mathbf{\Sigma}, \quad j = 1, \ldots, k, \tag{4.100}$$

onde $\boldsymbol{\mu}_j$ e $\mathbf{\Sigma}_j$ são como em (4.21) e (4.22), respectivamente. O estudo será baseado sob as condições de identificabilidade $\lambda_{e_j}$ ou $\lambda_{x_j}$, $j = 1, \ldots, k$, conhecidos e sob a parametrização ortogonal definida em (4.20). Note que as hipóteses nulas definidas em (4.99) e (4.100) são equivalentes a

$$H_0 : \beta_j = \beta, \phi_{3j} = \phi_3, \phi_{4j} = \phi_4, \quad j = 1, \ldots, k,$$

e

$$H_0 : \phi_{1j} = \phi_1, \phi_{2j} = \phi_2, \phi_{3j} = \phi_3, \phi_{4j} = \phi_4, \beta_j = \beta, \quad j = 1, \ldots, k,$$

respectivamente.

### 4.4.3 O teste da razão de verossimilhança para a hipótese $H_0$ : $\mathbf{\Sigma}_j = \mathbf{\Sigma}$, $j = 1, \ldots, k$

Sob normalidade, o logaritmo da função de verossimilhança $L(\boldsymbol{\phi})$ pode ser expressa como em (4.94). Assim, sob o modelo não restrito, temos que os EMV de $\boldsymbol{\mu}_j$ e $\mathbf{\Sigma}_j$ são dados como em (4.13) e substituindo em $L(\boldsymbol{\phi})$, temos que

$$L(\widehat{\boldsymbol{\phi}}) = \sum_{j=1}^{k} \left(-\frac{n_j}{2}\right) \log(\widehat{\phi}_{3j}\widehat{\phi}_{4j}) - n. \tag{4.101}$$

Por outro lado, os EMV de $\boldsymbol{\phi}_j$ sob a hipótese nula $H_0$ são dados por

$$\widetilde{\boldsymbol{\mu}}_j = \widehat{\boldsymbol{\mu}}_j = \boldsymbol{\mu}(\widehat{\boldsymbol{\phi}}_{L_j}), \quad \widetilde{\phi}_l = \overline{\boldsymbol{\alpha}}_l^{\top}(\widetilde{\beta}) S \overline{\boldsymbol{\alpha}}_l(\widetilde{\beta}), \quad l = 3, 4,$$

com $\widetilde{\beta}$ sendo a raiz da equação

$$\overline{\boldsymbol{\alpha}}_3^{\top}(\widetilde{\beta}) \mathbf{S} \overline{\boldsymbol{\alpha}}_4(\widetilde{\beta}) = 0,$$

onde $\mathbf{S}$ é como em (4.11) e $\mathbf{S} \sim W_2(n-k, \frac{1}{n}\mathbf{\Sigma})$, $n = n_1 + \cdots + n_k$, com $\mathbf{\Sigma} = \mathbf{\Sigma}(\boldsymbol{\phi}_E)$, $\boldsymbol{\phi}_E = (\phi_3, \phi_4, \beta)^{\top}$. Notemos que $\mathbf{S}$ pode ser expressa como

$$\mathbf{S} = \begin{pmatrix} S_{YY} & S_{XY} \\ S_{YX} & S_{XX} \end{pmatrix}, \tag{4.102}$$

onde

$$S_{YY} = \frac{1}{n}\sum_{j=1}^{k}\sum_{i=1}^{n_j}(Y_{ij} - \overline{Y}_j)^2, \quad S_{XY} = \frac{1}{n}\sum_{j=1}^{k}\sum_{i=1}^{n_j}(X_{ij} - \overline{X}_j)(Y_{ij} - \overline{Y}_j) \quad \text{e}$$

$$S_{XX} = \frac{1}{n}\sum_{j=1}^{k}\sum_{i=1}^{n_j}(X_{ij} - \overline{X}_j)^2,$$

o que implica que, sob $H_0$,

$$\widetilde{\phi}_3\widetilde{\phi}_4 = |\Sigma(\widetilde{\boldsymbol{\phi}}_E)| = S_{YY.X}S_{XX}, \tag{4.103}$$

onde $S_{YY.X} = S_{YY} - S_{XY}^2/S_{XX}$. Assim, de (4.100), e das propriedades da matriz Wishart (veja Muirhead, 1982), temos que

$$\frac{nS_{YY.X}}{\sigma_{YY.X}} \sim \chi^2_{n-k-1} \quad \text{e} \quad n\frac{S_{XX}}{\sigma_{XX}} \sim \chi^2_{n-k} \tag{4.104}$$

são independentes, onde $\sigma_{YY.X} = \sigma_{YY} - \sigma_{XY}^2/\sigma_{XX}$. Além disso, sob $H_0$, temos de (4.103) e (4.104) que

$$\begin{aligned} E_0\left[\log\left(\frac{\widetilde{\phi}_3\widetilde{\phi}_4}{\phi_3\phi_4}\right)\right] &= E_0\left[\log\left(\frac{S_{YY.X}S_{XX}}{\sigma_{YY.X}\sigma_{XX}}\right)\right] \\ &= E_0\left[\log\left(\frac{S_{YY.X}}{\sigma_{YY.X}}\right)\right] + E_0\left[\log\left(\frac{S_{XX}}{\sigma_{XX}}\right)\right] \\ &= \psi\left(\frac{n-k-1}{2}\right) + \psi\left(\frac{n-k}{2}\right) - 2\log\frac{n}{2}, \end{aligned} \tag{4.105}$$

onde na última igualdade usamos o fato de que se $V \sim G_a(a,b)$, então $E(\log V) = \psi(a) - \log b$, com $\psi$ sendo a função digama definida na seção 4. É fácil ver que neste caso a estatística da razão de verossimilhança $G$ tem distribuição assintótica $\chi^2_{3(k-1)}$.

**Teorema 4.2** *Sob o modelo normal e a condição de identificabilidade $\lambda_{e_j}$ ou $\lambda_{xj}$, $j = 1,\ldots,k$, conhecidos, a estatística da razão de verossimilhança $G$ e sua estatística corrigida $G^*$ para testar $H_0$ são dadas por*

$$G = \sum_{j=1}^{k} n_j \log\left(\frac{\widetilde{\phi}_3\widetilde{\phi}_4}{\widehat{\phi}_{3j}\widehat{\phi}_{4j}}\right) \quad e \quad G^* = (1+d)^{-1}G, \tag{4.106}$$

*respectivamente, onde*

$$d = \frac{1}{18(k-1)}\left[\sum_{j=1}^{k}\frac{25}{n_j} - \frac{6(k+1)(k+2)+1}{n}\right], \quad n = n_1 + \cdots + n_k. \tag{4.107}$$

**Prova:** De (4.103) temos que sob a hipótese nula $H_0$

$$
\begin{aligned}
L(\widetilde{\phi}) &= \sum_{j=1}^{k}\left(-\frac{n_j}{2}\right)\log(\widetilde{\phi}_3\widetilde{\phi}_4) - \frac{n}{2}\mathrm{tr}(\Sigma^{-1}(\widetilde{\phi}_E)\mathbf{S}) \qquad (4.108)\\
&= \sum_{j=1}^{k}\left(-\frac{n_j}{2}\right)\log(\widetilde{\phi}_3\widetilde{\phi}_4) - n.
\end{aligned}
$$

Assim, de (4.101) e (4.108), temos a primeira parte da demonstração.

Por outro lado, de (4.105) e (4.83), temos sob a hipótese nula $H_0$ que

$$
\begin{aligned}
E_0[G] &= \sum_{j=1}^{k} n_j\left\{E_0\left[\log\left(\frac{\widetilde{\phi}_3\widetilde{\phi}_4}{\phi_3\phi_4}\right)\right] - E_0\left[\log\left(\frac{\widehat{\phi}_{3j}\widehat{\phi}_{4j}}{\phi_{3j}\phi_{4j}}\right)\right]\right\}\\
&= n\left[\psi\left(\frac{n-k-1}{2}\right) + \psi\left(\frac{n-k}{2}\right) - 2\log\frac{n}{2}\right]\\
&\quad -\sum_{j=1}^{k} n_j\left[\psi\left(\frac{n_j-2}{2}\right) + \psi\left(\frac{n_j-1}{2}\right) - 2\log\frac{n_j}{2}\right].
\end{aligned}
$$

Usando o Lema 4.1, obtemos que

$$
\begin{aligned}
&n_j\left[\psi\left(\frac{n_j-2}{2}\right) + \psi\left(\frac{n_j-1}{2}\right) - 2\log\frac{n_j}{2}\right]\\
&\quad = n_j\left\{\left[\psi\left(\frac{n_j-1}{2}\right) - \psi\left(\frac{n_j-2}{2}\right)\right] - 2\left[\psi\left(\frac{n_j-1}{2}\right) - \log\frac{n_j}{2}\right]\right\}\\
&\quad = \left(1 + \frac{5}{2n_j} + O(n_j^{-2}) + 4 + \frac{5}{3n_j} + O(n_j^{-2})\right)\\
&\quad = 5 + \frac{25}{6n_j} + O(n_j^{-2}). \qquad (4.109)
\end{aligned}
$$

Assim, de (4.77) e (4.109) juntamente com ii) do Lema 4.1, temos que

$$
\begin{aligned}
E_0[G] &= -2k - 3 - \frac{6(k+1)(k+2)+1}{6n} + O(n^{-2}) + 5k + \sum_{j=1}^{k}\frac{25}{6n_j} + O(n_*^{-2})\\
&= 3(k-1) + \frac{1}{6}\left(\sum_{j=1}^{k}\frac{25}{n_j} - \frac{6(k+1)(k+2)+1}{n}\right) + O(n_*^{-2}).
\end{aligned}
$$

Portanto, a estatística $G^* = (1+d)^{-1}G$ tem distribuição mais próxima da distribuição $\chi^2_{3(k-1)}$ que a distribuição de $G$. □

Seja $G_E$ a estatística da razão de verossimilhança sob o modelo elíptico (como na Seção 4.4.2), os resultados seguintes são enunciados para o modelo elíptico.

**Corolário 4.3** *Sob o modelo elíptico correspondente à especificação (A) e supondo que $\lambda_{e_j}$ ou $\lambda_{x_j}$, $j = 1, \ldots, k$, conhecidos, a estatística corrigida de $G_E$ é dada por*

$$G_E^* = (1 + d)^{-1} G_E, \tag{4.110}$$

*onde $d$ é como no Teorema 4.2.*

**Prova:** Sob a hipótese não restrita, temos que $L(\widehat{\phi})$ é dado por (4.85). Além disso, sob a hipótese nula, temos que

$$\widetilde{\boldsymbol{\mu}}_j = \widehat{\boldsymbol{\mu}}_j = \overline{\mathbf{Z}}_j \quad \text{e} \quad \widetilde{\Sigma} = \boldsymbol{\Sigma}(\widetilde{\boldsymbol{\phi}}_E) = \frac{np}{u_f} \widetilde{\Sigma}_N,$$

onde $\widetilde{\boldsymbol{\Sigma}}_N = \mathbf{S}$ (definido em (4.11)); logo, substituindo em (4.84) segue que

$$L(\widetilde{\phi}) = \sum_{j=1}^{k} \left(-\frac{n_j}{2}\right) \log \left| \frac{np}{u_f} \boldsymbol{\Sigma}(\widetilde{\phi}_E) \right| + \log f(u_f). \tag{4.111}$$

Assim,

$$G_E = \sum_{j=1}^{k} n_j \log \left( \frac{\widetilde{\phi}_3 \widetilde{\phi}_4}{\widehat{\phi}_{3j} \widehat{\phi}_{4j}} \right),$$

onde $\widetilde{\phi}_3$, $\widetilde{\phi}_4$, $\widehat{\phi}_{3j}$ e $\widehat{\phi}_{4j}$, $j = 1, \ldots, k$, são os EMV sob o modelo normal. Assim, pelo Teorema 4.2, temos que $G_E^* = (1 + d)^{-1} G_E$, onde $d$ é como em (4.107). □

### 4.4.4 O teste da razão de verossimilhança para a hipótese $H_0 : \boldsymbol{\mu}_j = \boldsymbol{\mu}$, $\boldsymbol{\Sigma}_j = \boldsymbol{\Sigma}$, $j = 1, \ldots, k$

Como na seção anterior, trabalharemos sob a parametrização ortogonal definida em (4.20) e com as condições de idetificabilidade definidas em (4.14). Nesta seção estamos interessados em testar a hipótese nula $H_0 : \boldsymbol{\mu}_j = \boldsymbol{\mu}$, $\boldsymbol{\Sigma}_j = \boldsymbol{\Sigma}$, $j = 1, \ldots, k$, usando a estatística da razão de verossimilhança $G$ que neste caso tem distribuição assintótica $\chi^2_{5(k-1)}$. Com efeito, de (4.27), os EMV sob $H_1$ de $\boldsymbol{\phi}_j$, $j = 1, \ldots, k$, são dados por (4.28), (4.29) e (4.30), e os EMV sob $H_0$ são dados por

$$\widetilde{\boldsymbol{\mu}} = \overline{\mathbf{Z}} = \frac{1}{n} \sum_{j=1}^{k} n_j \overline{\mathbf{Z}}_j, \quad n = n_1 + \cdots + n_k, \tag{4.112}$$

$$\widetilde{\phi}_l = \overline{\boldsymbol{\alpha}}_l^\top(\widetilde{\beta}) \mathbf{S}_{(n)} \overline{\boldsymbol{\alpha}}_l(\widetilde{\beta}), \quad l = 3, 4, \tag{4.113}$$

onde

$$\mathbf{S}_{(n)} = \frac{1}{n}\sum_{j=1}^{k}\sum_{i=1}^{n_j}(\mathbf{Z}_{ij}-\overline{\mathbf{Z}})(\mathbf{Z}_{ij}-\overline{\mathbf{Z}}) = \begin{pmatrix} S_{(n)YY} & S_{(n)XY} \\ S_{(n)YX} & S_{(n)XX} \end{pmatrix}, \tag{4.114}$$

com

$$S_{(n)YY} = \frac{1}{n}\sum_{j=1}^{k}\sum_{i=1}^{n_j}(Y_{ij}-\overline{Y})^2, \quad S_{(n)XY} = \frac{1}{n}\sum_{j=1}^{k}\sum_{i=1}^{n_j}(X_{ij}-\overline{X})(Y_{ij}-\overline{Y})$$

e

$$S_{(n)XX} = \frac{1}{n}\sum_{j=1}^{k}\sum_{i=1}^{n_j}(X_{ij}-\overline{X})^2,$$

e o EMV de $\beta$ é raiz da equação $\overline{\boldsymbol{\alpha}}_3^\top(\widetilde{\beta})\mathbf{S}_{(n)}\overline{\boldsymbol{\alpha}}_4(\widetilde{\beta}) = 0$. Note que de (4.113), (4.114), juntamente com o fato que $|\boldsymbol{\Sigma}(\boldsymbol{\phi}_E)| = \phi_3\phi_4$, temos que $\widetilde{\phi}_3\widetilde{\phi}_4 = |\boldsymbol{\Sigma}(\widetilde{\boldsymbol{\phi}}_E)| = S_{(n)YY.X}S_{(n)XX}$, onde $S_{(n)YY} = S_{(n)YY} - S^2_{(n)XY}/S_{(n)XX}$. Conseqüentemente,

$$L(\widetilde{\boldsymbol{\phi}}) = \sum_{j=1}^{n} n_j \log(\widetilde{\phi}_3\widetilde{\phi}_4) - n\ . \tag{4.115}$$

Além disso, sob $H_0$, $\overline{\mathbf{Z}} \sim N_2(\boldsymbol{\mu},\boldsymbol{\Sigma})$ e $\mathbf{S}_{(n)} \sim W_2(n-1,\frac{1}{n}\boldsymbol{\Sigma})$, de onde, pela propriedade da matriz Wishart, segue que

$$\frac{nS_{(n)YY.X}}{\sigma_{YY.X}} \sim \chi^2_{n-2} \quad \text{e} \quad \frac{nS_{(n)XX}}{\sigma_{XX}} \sim \chi^2_{n-1}\ . \tag{4.116}$$

**Teorema 4.3** *Sob o modelo normal e a condição de identificabilidade $\lambda_{e_j}$ ou $\lambda_{x_j}$, $j = 1,\ldots,k$, conhecidos, a estatística da razão de verossimilhança $G$ e sua estatística modificada $G^*$ do teste $H_0$ são dadas por*

$$G = \sum_{j=1}^{k} n_j \log\left(\frac{\widetilde{\phi}_3\widetilde{\phi}_4}{\widehat{\phi}_{3j}\widehat{\phi}_{4j}}\right) \quad e \quad G_* = (1+d)^{-1}G, \tag{4.117}$$

*respectivamente, onde*

$$d = \frac{1}{30(k-1)}\left[\sum_{j=1}^{k}\frac{25}{n_j} - \frac{37}{n}\right]. \tag{4.118}$$

**Prova:** De (4.101) e (4.115) temos a expressão da estatística da razão de verossimilhança $G$. Agora, considerando (4.116) a definição da função digama $\psi$, a esperança de $G$ sob

$H_0$ é dada por

$$\begin{aligned} E_0[G] &= \sum_{j=1}^{k} n_j \left\{ E_0 \left[ \log \left( \frac{\widetilde{\phi}_3 \widetilde{\phi}_4}{\phi_3 \phi_4} \right) \right] - E_0 \left[ \log \left( \frac{\widehat{\phi}_{3j} \widetilde{\phi}_{4j}}{\phi_{3j} \phi_{4j}} \right) \right] \right\} \\ &= n \left[ \psi \left( \frac{n-2}{2} \right) + \psi \left( \frac{n-1}{2} \right) - 2 \log \frac{n}{2} \right] \\ &\quad - \sum_{j=1}^{k} n_j \left[ \psi \left( \frac{n_j - 2}{2} \right) + \psi \left( \frac{n_j - 1}{2} \right) - 2 \log \frac{n_j}{2} \right]. \end{aligned}$$

Das propriedades da função digama em (4.74), (4.75) e (4.76) (ou do Lema 4.1, com $k = 1$), temos que

$$n \left[ \psi \left( \frac{n-2}{2} \right) + \psi \left( \frac{n-1}{2} \right) - 2 \log \frac{n}{2} \right] = -5 - \frac{37}{6n}.$$

Assim, da igualdade acima e de (4.109), temos que

$$\begin{aligned} E_0[G] &= -5 - \frac{37}{6n} + O(n^{-2}) + 5d + \sum_{j=1}^{k} \frac{25}{6n_j} + O(n_*^{-2}) \\ &= 5(k-1) + \frac{1}{6} \left[ \sum_{j=1}^{k} \frac{25}{n_j} - \frac{37}{n} \right] + O(n_*^{-2}). \end{aligned}$$

Assim, $G^* = (1+d)^{-1}$ tem distribuição mais próxima da distribuição $\chi^2_{5(k-1)}$ que a estatística $G$. □

**Corolário 4.4** *Sob o modelo elíptico correspondente à especificação (A) e $\lambda_{e_j}$ ou $\lambda_{x_j}$, $j = 1, \ldots, k$, conhecidos, a estatística corrigida $G_E$ é dada por*

$$G_E^* = (1+d)^{-1} G_E,$$

*com $d$ como no Teorema 4.3.*

**Prova:** Os EMV $\widehat{\boldsymbol{\mu}}_j$ e $\widehat{\boldsymbol{\Sigma}}_j$, $j = 1, \ldots, k$, são como em (4.53). Assim, $L(\widehat{\boldsymbol{\phi}})$ é dado como (4.85). Por outro lado, os EMV sob $H_0$ são dados por

$$\widetilde{\boldsymbol{\mu}} = \overline{\mathbf{Z}} \quad \text{e} \quad \widetilde{\boldsymbol{\Sigma}}_j = \frac{np}{u_f} \widetilde{\boldsymbol{\Sigma}}_{N_j} = \frac{np}{u_f} \mathbf{S}_{(n)}.$$

Assim,

$$L(\widetilde{\phi}) = \sum_{j=1}^{k} \left(-\frac{n_j}{2}\right) \log \left|\frac{np}{u_f} \mathbf{S}_{(n)}\right| + \log f(u_f). \tag{4.119}$$

Assim, de (4.85) e (4.114), segue que

$$G_E = \sum_{k=1}^{k} n_j \log \left(\frac{\widetilde{\phi}_3 \widetilde{\phi}_4}{\widehat{\phi}_{3j} \widehat{\phi}_4}\right)$$

e pelo Teorema 4.3 temos que $G_E^* = (1+d)^{-1} G_E$. □

Note que para o modelo elíptico sob a especificação (C) não é possível aplicar a metodologia usada sob o modelo normal ou sob o modelo elíptico seguindo a especificação (A), para obter o fator de correção de Bartlett. Neste caso, sugerimos usar o método de Lawley (1976), para obter o fator de Bartlett. Note também que no modelo elíptico sob a especificação (B) nem sempre é possível obter diretamente a $E_0[G]$, como por exemplo, quando testamos a hipótese nula $H_0 : \mathbf{\Sigma}_j = \mathbf{\Sigma}$, $j = 1, \ldots, k$. Um método alternativo é usar o método de Lawley (1976).

# Capítulo 5

# Conclusões

No capítulo 2 desta tese, estudamos a consistência e a distribuição assintótica da matriz de covariâncias amostrais $\mathbf{S}_n$ de $n$ vetores aleatórios $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$, independentes e de dimensão $p$, com vetores de média $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_n$, respectivamente, e com matrizes de covariâncias iguais a $\Sigma$; com distribuição não especificada tendo quartos momentos finitos. Estes resultados são aplicados ao modelo ultraestrutural elíptico para estudar a consistência e distribuição assintótica do estimador obtido pelo método dos momentos e do estimador obtido através do método de Mínimos Quadrados Generalizados (GLS). As condições consideradas por Gleser (1985) para os parâmetros incidentais $\mu_{x_1}, \ldots, \mu_{x_n}$ sob o modelo ultraestrutural normal são a existência de $\mu_x$ e $\sigma^*_{xx} > 0$ tais que

$$\sqrt{n}(\overline{\mu}_x - \mu_x) \to 0 \quad \text{e} \quad \sqrt{n}(S^*_{xx} - \sigma^*_{xx}) \to 0 \ ,$$

onde $\overline{\mu}_x$ e $S^*_{xx}$ são como na condição S2. No entanto, acreditamos que é necessário condições adicionais para garantir a convergência em distribuição do vetor $\mathrm{Vec}(\mathbf{S}_n)$, pois a prova do corolário 4.1 em Gleser (1981) utilizado na prova da convergência em distribuição de $\mathbf{S}_n$ não é válida. Para garantir a convergência em distribuição, consideramos que os parâmetros $\mu_{x_1}, \ldots, \mu_{x_n}$ satisfazem a condição adicional (veja seção 2.3 do capítulo 2)

$$\max_{1 \le k \le n} \frac{(\mu_{x_k} - \mu_x)^2}{n} \to 0 \ .$$

Em particular, nos modelos estrutural e funcional elíptico, estudamos a eficiência relativa assintótica do estimador de máxima verossimilhança com respeito aos estimadores estudados usando a distribuição assintótica da matriz de covariâncias amostrais $\mathbf{S}_n$.

No capítulo 3, baseado no trabalho de Mak (1982), estudamos a consistência e a distribuição assintótica do estimador de máxima verossimilhança no modelo funcional multiunivariado, quando o vetor de erros tem uma distribuição elíptica com vetor de locações zero e matriz de dispersão identidade. Para o caso univariado, estudamos a eficiência relativa assintótica do estimador de máxima verossimilhança com respeito ao estimador de Mínimos Quadrados Generalizados. Um trabalho para futura pesquisa é fazer inferência no modelo funcional quando o vetor de erros tem uma distribuição elíptica com vetor de locação zero e matriz de dispersão $\sigma^2\mathbf{I}_2$, com $\sigma^2$ desconhecido. Sob normalidade, este caso é estudado em Mak (1982). Outro trabalho de pesquisa a ser desenvolvido é o estudo de testes de hipóteses para testar hipóteses do tipo $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$, com $\boldsymbol{\theta}_0$ especificado. Tal estudo pode ser baseado na estatística do tipo Wald dada por

$$W = n(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^\top \widetilde{\boldsymbol{\Sigma}}_n^{-1} (\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0),$$

onde $\boldsymbol{\Sigma}_n$ é a matriz de covariâncias assintótica do estimador de máxima verossimilhança $\widetilde{\boldsymbol{\theta}}_n$. Do teorema 3.6 pode-se mostrar que $W$ tem distribuição assintótica $\chi^2_{2q}$. Outra estatística que pode ser usada para testar $H_0$ é a estatística dada por

$$S = nU_n^\top(\boldsymbol{\theta}_0)\widetilde{V}_n^{-1}(\boldsymbol{\theta}_0)U_n(\boldsymbol{\theta}_0),$$

onde

$$U_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial}{\partial\theta_1}\log(t_i^2), \ldots, \frac{\partial}{\partial\theta_{2q}}\log f(t_i^2)\right)$$

pois, usando a demonstração do Teorema 2.3 em Mak (1982), pode ser mostrado que

$$\sqrt{n}V_n^{-1/2}(\boldsymbol{\theta})U_n(\boldsymbol{\theta}) \xrightarrow{\mathrm{d}} N_{2q}(\mathbf{0}, \mathbf{I}_{2q}),$$

o que implica que $S$ tem distribuição assintótica $\chi^2_{2q}$.

No capítulo 4, discutimos a inferência sobre os coeficientes angulares $\beta_1, \ldots, \beta_k$ do modelo consistindo de duas ou mais populações estruturais, usando a estatística da razão de verossimilhança. Obtemos o fator de Bartlett sob o modelo normal, através da avaliação direta do valor esperado da estatística da razão de verossimilhança sob a hipótese nula, que representa um método diferente daquele introduzido por Lawley (1956). Neste enfoque o fator de correção depende de funções de cumulantes conjuntos de derivadas até quarta ordem do logaritmo da função de verossimilhança que em situações particulares envolvem manipulações algébricas muito trabalhosas. Uma contribuição importante que apresentamos é a extensão de alguns resultados obtidos sob normalidade para a classe das distribuições elípticas sob três especificações diferentes que denotamos por (A), (B) e (C). Notamos que em situações em que não é possível obter-se diretamente o valor esperado $E_0[G]$, como acontece, por exemplo, no modelo elíptico sob a especificação (C), sugerimos usar o procedimento de Lawley (1956) como foi desenvolvido por Arellano-Valle (1994) sob o modelo estrutural elíptico univariado.

# Apêndice A

# Distribuições esféricas e elípticas

Neste apêndice apresentamos algumas propriedades das distribuições elípticas multivariadas, que serão de utilidade na demonstração dos resultados apresentados nos capítulos anteriores.

As distribuições esféricas são caracterizadas por sua invariância com respeito a transformações ortogonais e formam uma classe geral de distribuições com a mesma simetria esférica que a distribuição normal padrão. As distribuições elípticas, entretanto, podem ser geradas mediante transformações (lineares) de locação e escala das distribuições esféricas. Conseqüentemente, as distribuições elípticas constituem uma classe bem geral de famílias paramétricas de distribuições probabilísticas que preservam a estrutura simétrica da família de distribuições normais. De fato, a família das distribuições normais é um elemento particular desta classe. Além da família de distribuições normais, outros elementos típicos são as famílias normal "contaminada", $t$-Student e outras que veremos na próxima seção. Contudo, existe um grande número de modelos paramétricos alternativos ao modelo normal, que podem ser obtidos a partir da classe das distribuições elípticas.

Embora o interesse por estas distribuições seja relativamente recente, muitas de suas propriedades têm sido motivo de estudo por diferentes autores; o primeiro desenvolvimento teórico bem organizado é devido a Kelker (1970). Desde então, estas distribuições têm sido o objetivo de estudo em termos de pesquisa teórica e em aplicações em diferentes áreas da estatística. Resultados recentes sobre as propriedades das distribuições elípticas, podem ser encontrados, entre outros, em Mitchell e Kazano (1985), Berkane e Bentler (1986, 1987), Muirhead (1980, 1982), Mitchel (1988, 1989) e Fang, Kotz e Ng (1990). Destaca-se ainda o excelente trabalho de Cambanis, Huang e Simons (1981), pelo tratamento rigoroso e sistemático das distribuições elípticas.

## A.1 Definições e exemplos

No que segue desta seção, $\mathbf{X}$ e $\mathbf{Y}$ são vetores aleatórios $n$-dimensionais e o conjunto

$$\mathbf{O}_n = \{\mathbf{\Gamma} : n \times n;\ \mathbf{\Gamma}^{\top}\mathbf{\Gamma} = \mathbf{\Gamma}\mathbf{\Gamma}^{\top} = I_n\} \tag{A.1}$$

denotará o grupo ortogonal das matrizes $n \times n$.

**Definição A.1** *Um vetor aleatório $\mathbf{X} : n \times 1$ é dito ter uma distribuição esférica se para todo $\mathbf{\Gamma} \in O_n$ tem-se que*

$$\mathbf{X} \stackrel{\mathrm{d}}{=} \mathbf{\Gamma}\mathbf{X}. \tag{A.2}$$

*Equivalentemente, esta distribuição pode ser definida em termos da função característica, isto é, se $\phi_{\mathbf{X}}(\mathbf{t}) = Ee^{i\mathbf{t}^{\top}\mathbf{X}}$, $\mathbf{t} \in \mathbb{R}^n$, então*

$$\phi_{\mathbf{X}}(\mathbf{t}) = \phi(\mathbf{t}^{\top}\mathbf{t}), \tag{A.3}$$

*para alguma função $\phi(u)$, $u \geq 0$. A distribuição esférica também pode ser definida em termos da função densidade (se $\mathbf{X}$ tiver densidade), ou seja, $\mathbf{X}$ tem distribuição esférica se a função densidade é da forma*

$$P_{\mathbf{X}}(\mathbf{x}) = f(\mathbf{x}^{\top}\mathbf{x}),\ \ \mathbf{x} \in \mathbb{R}^n, \tag{A.4}$$

*para alguma função não negativa $f$ tal que $r^{n-1}f(r^2)$ é integrável sobre $[0, \infty)$, que garante que $P_{\mathbf{X}}(\mathbf{x})$ seja uma função densidade.*

A seguir apresentamos alguns exemplos de distribuições esféricas tomadas em Fang et al. (1990), onde $f(\mathrm{x}^\top \mathrm{x})$, $\mathrm{x} \in \mathbb{R}^n$, representa a função densidade e $\phi(\mathbf{t}^\top \mathbf{t})$, $\mathbf{t} \in \mathbb{R}^n$, a função característica de um vetor $\mathbf{X}$ que tem distribuição esférica.

**Tabela A.1:** Subclasse de distribuições esféricas $n$-variadas

| Tipo | $f(u)$; $u = \mathrm{x}^\top \mathrm{x}$, $\mathrm{x} \in \mathbb{R}^n$ ou $\phi(u)$; $u = \mathbf{t}^\top \mathbf{t}$, $\mathbf{t} \in \mathbb{R}^n$ |
|---|---|
| Kotz | $f(u) = cu^{N-1} \exp(-ru^S)$; $r, s > 0$, $2N + n > 2$ |
| Normal | $f(u) = c\exp(-u/2)$ |
| Pearson VII | $f(u) = c(1 + u/S)^{-N}$; $N > n/2$, $S > 0$ |
| $t$ generalizada | $f(u) = c(1 + u/S)^{-(n+m)/2}$; $S, m > 0$ ($m$ inteiro) |
| Cauchy | $f(u) = c(1 + u/S)^{-(n+1)/2}$; $S > 0$ |
| Pearson II | $f(u) = c(1 - u)^m$; $m > 0$, $0 \leq u < 1$ |
| Logística | $f(u) = c\exp(-u)/(1 + \exp(-u))^2$ |
| Bessel | $f(u) = c(\sqrt{u}/\beta)^a Ka(\sqrt{u}/\beta)$; $a > -n/2$, $\beta > 0$, onde $Ka(\cdot)$ denota a função de Bessel modificada de terceira ordem |
| Normal composta | $f(u) = c\int_0^\infty v^{-n/2} \exp(-u/2v) \mathrm{d}F(u)$; $F$ é uma F.D. |
| Lei estável | $\phi(u) = \exp(ru^{\alpha/2})$; $0 < \alpha \leq 2$, $r < 2$ |
| Uniforme | $\phi(u) = {}_0F_1(n/2, -u^2/4)$; onde ${}_0F_1(\cdot)$ é a função hipergeométrica generalizada |

**Definição A.2** *Um vetor aleatório* $\mathbf{X} : n \times 1$ *é dito ter distribuição elíptica (n-variada) com vetor de locação* $\boldsymbol{\mu} : n \times 1$ *e matriz de dispersão* $\boldsymbol{\Sigma} : n \times n$ *não negativa se a função característica de* $\mathbf{X}$ *é da forma*

$$\phi_{\mathbf{X}}(t) = e^{i\mathbf{t}^\top \boldsymbol{\mu}} \phi(\mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t}), \quad \mathbf{t} \in \mathbb{R}^n, \tag{A.5}$$

*para alguma função* $\phi$. *Em tal caso dizemos que* $\mathbf{X}$ *tem distribuição elíptica n-variada (ou simplesmente elíptica) com parâmetros* $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ *e* $\phi$, *e escrevemos*

$$\mathbf{X} \sim E\ell_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad ou \quad \mathbf{X} \sim E\ell_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \phi), \tag{A.6}$$

*quando é necessário indicar a dependência de* $\phi$ *na representação paramétrica.*

Outra definição equivalente de distribuição elíptica é $\mathbf{X} \sim E\ell(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \phi)$, se

$$\mathbf{X} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + \mathbf{A}\mathbf{Z}, \tag{A.7}$$

onde $\mathbf{A}$ é uma matriz $n \times k$, com posto$(\mathbf{A}) = k$, $\mathbf{A}\mathbf{A}^\top = \boldsymbol{\Sigma}$ e o vetor aleatório $\mathbf{Z}$ tem distribuição esférica $k$-dimensional com função característica $\phi(\mathbf{t}^\top \mathbf{t})$, $\mathbf{t} \in \mathbb{R}^n$ (veja Fang et al., 1990).

Se a distribuição de $\mathbf{X}$ tiver densidade (com respeito à medida de Lebesgue em $\mathbb{R}^n$), então $\boldsymbol{\Sigma} > 0$ (Arnold, 1981) e sua densidade é dada por

$$P_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-1/2} f((\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})), \ \mathbf{x} \in \mathbb{R}^n, \tag{A.8}$$

para alguma função não negativa $f(u)$, $u \geq 0$. Neste caso, costuma-se substituir $\phi$ por $f$ na notação dada em (A.6). Observemos que se $\mathbf{X} \sim E\ell_n(\mathbf{0}, \mathbf{I}_n; \phi)$, então, de (A.5) e (A.7), temos que a distribuição de $\mathbf{X}$ é esférica. Assim, a classe das distribuições esféricas é uma subclasse da classe das distribuições elípticas.

Seja $\mathbf{U}^{(n)} : n \times 1$ um vetor aleatório com distribuição uniforme sobre o conjunto

$$S_n = \{\mathbf{x} \in \mathbb{R}^n; \|\mathbf{x}\| = 1\|\}, \tag{A.9}$$

a superfície da esfera unitária de $\mathbb{R}^n$, isto é,

$$\mathbf{U}^{(n)} \sim U(S_n),$$

Kariya e Eaton (1977) e Muirhead (1982) mostram que a distribuição uniforme é a única distribuição sobre $S_n$ que é invariante por transformações ortogonais ($\mathbf{U}^{(n)} \stackrel{\mathrm{d}}{=} \boldsymbol{\Gamma}\mathbf{U}^{(n)}$, para todo $\boldsymbol{\Gamma} \in O_n$) e uma das propriedades importantes é que ela permite caracterizar as distribuições esféricas, como veremos nos resultados seguintes.

## A.2 Propriedades

**Teorema A.1** *Seja $\mathbf{X}$ um vetor aleatório n-dimensional. Então $\mathbf{X} \stackrel{\mathrm{d}}{=} \boldsymbol{\Gamma}\mathbf{X}$ se, e somente se, $\mathbf{X} \stackrel{\mathrm{d}}{=} R\mathbf{U}^{(n)}$, para alguma variável aleatória $R \geq 0$ independente de $\mathbf{U}^{(n)}$. Além disso, se $\mathbf{X} \sim E\ell_n(\mathbf{0}, \mathbf{I}_n; \phi)$, $P(\mathbf{X} = \mathbf{0}) = 0$, então*

$$\frac{\mathbf{X}}{\|\mathbf{X}\|} \stackrel{\mathrm{d}}{=} \mathbf{U}^{(n)} \tag{A.10}$$

*e é independente de* $\|\mathbf{X}\| \overset{\mathrm{d}}{=} R$.

A demonstração do teorema acima pode ser encontrada em Kariya e Eaton (1977) e Fang et al. (1990).

Para $\mathbf{X} \sim E\ell_p(\mathbf{0}, \mathbf{I}_p; f)$, isto é, $\mathbf{X}$ tem densidade como em (A.4), e $\mathbf{A}$ e $\mathbf{B}$ matrizes simétricas de dimensão $p \times p$. Seja também $\mathbf{U} = \mathbf{U}^{(n)}$. Então, Mitchell (1989) mostrou que

$$E[\mathbf{U}^\top \mathbf{A}\mathbf{U}] = \frac{1}{p}\mathrm{tr}(\mathbf{A}) \tag{A.11}$$

e

$$E[(\mathbf{U}^\top \mathbf{A}\mathbf{U})(\mathbf{U}^\top \mathbf{B}\mathbf{U})] = \frac{1}{p(p+2)}\{2\mathrm{tr}(\mathbf{AB}) + \mathrm{tr}(\mathbf{A})\mathrm{tr}(\mathbf{B})\}. \tag{A.12}$$

Além disso, se $W_f(u) = f'(u)/f(u)$, então desde que os momentos envolvidos existam, tem-se que

$$E[(\mathbf{X}^\top \mathbf{A}\mathbf{X})(W_f(\|\mathbf{X}\|^2))^m] = \frac{1}{p}\mathrm{tr}(\mathbf{A})E[\|\mathbf{X}\|^2(W_f(\|\mathbf{X}\|^2))^m] \tag{A.13}$$

e

$$\begin{aligned} &E[(\mathbf{X}^\top A\mathbf{X})(\mathbf{X}^\top \mathbf{B}\mathbf{X})(W_f(\|\mathbf{X}\|^2))^m] \\ &= \frac{1}{p(p+2)}\{2\mathrm{tr}(\mathbf{AB}) + \mathrm{tr}(\mathbf{A})\mathrm{tr}(\mathbf{B})\}E[\|\mathbf{X}\|^4(W_f(\|\mathbf{X}\|^2)^m)]. \end{aligned} \tag{A.14}$$

**Teorema A.2** *Seja* $\mathbf{X} \sim E\ell(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \phi)$. *Então, desde que existam*

$$E[\mathbf{X}] = \boldsymbol{\mu} \quad e \quad \mathrm{Cov}[\mathbf{X}] = \delta\boldsymbol{\Sigma}, \tag{A.15}$$

*onde* $\delta = -2\phi'(0)$ *e* $\phi'(0) = \left(\frac{\mathrm{d}}{\mathrm{d}u}\phi(u)\right)|_{u=0}$.

Por exemplo, se $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, então $\delta = 1$ e se $\mathbf{X} \sim t(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu)$, $\nu > 2$, então $\delta = \frac{\nu}{\nu - 2}$. Kelker (1970) nota que se $\mathbf{X} \sim E\ell_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}; f)$, então as componentes de $\mathbf{X}$ têm $k$ momentos finitos se, e somente se, $r^{k+n-1}f(r^2)$ é integrável sobre $[0, \infty)$. Os resultados seguintes estabelecem que a distribuição de qualquer transformação linear de um vetor aleatório com distribuição elíptica é também elíptica e algumas propriedades da distribuição condicional

**Teorema A.3** *Seja* $\mathbf{X} \sim E\ell_n(\boldsymbol{\mu}, \Sigma; \phi)$, $\eta \in \mathbb{R}^n$ *e* $\mathbf{A}$ *uma matriz* $n \times n$. *Então,*

$$\eta + \mathbf{A}\mathbf{X} \sim E\ell_m(\eta + \mathbf{A}\boldsymbol{\mu}, \mathbf{A}\Sigma\mathbf{A}^\top; \phi). \tag{A.16}$$

Se partirmos $\mathbf{X}$, $\boldsymbol{\mu}$ e $\Sigma$ como

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \quad \text{e} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \tag{A.17}$$

onde $\mathbf{X}_1$ e $\boldsymbol{\mu}_1$ são vetores $m$-dimensionais e $\boldsymbol{\Sigma}_{11}$ é uma matriz $m \times m$, com $1 \leq m \leq n$, então do teorema acima temos que

$$\mathbf{X}_1 \sim E\ell_m(\boldsymbol{\mu}_1, \Sigma_{11}; \phi) \quad \text{e} \quad \mathbf{X}_2 \sim E\ell_{n-m}(\boldsymbol{\mu}_2; \Sigma_{22}; \phi). \tag{A.18}$$

Para o teorema seguinte, vamos supor que $\mathbf{X}$ tem densidade $f(\mathbf{x}^\top\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^n$, e está partido como em (A.17).

**Teorema A.4** *Seja* $\mathbf{X} \sim E\ell_n(\boldsymbol{\mu}, \Sigma; f)$. *Então,*

*i)* $(\mathbf{X}_1 | \mathbf{X}_2 = \mathbf{x}_2) \sim E\ell_m(\boldsymbol{\mu}_1(\mathbf{x}_2), \Sigma_{11.2}; f_q(\mathbf{x}_2))$, *onde*

$$\begin{aligned}
&\boldsymbol{\mu}_1(\mathbf{x}_2) = \boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2), \quad \Sigma_{11.2} = \Sigma_{11} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12},\\
&q(\mathbf{x}_2) = (\mathbf{x}_2 - \boldsymbol{\mu}_2)^\top \Sigma_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2) \quad e\\
&f_a(u) = \frac{f(u+a)}{s_m}\left(\int_0^\infty f(\sigma^2 + a) r^{m-1} \mathrm{d}x\right)^{-1}, \quad u \geq 0;
\end{aligned}$$

*ii) assumindo que existem*

$$\begin{aligned}
&E[\mathbf{X}_1 | \mathbf{X}_2 = \mathbf{x}_2] = \boldsymbol{\mu}_2(\mathbf{x}_2), \quad \mathrm{Cov}[\mathbf{X}_1 | \mathbf{X}_2 = \mathbf{x}_2] = \alpha_{q(\mathbf{x}_2)}\boldsymbol{\Sigma}_{11.2}\\
e \qquad &\phi_{\mathbf{X}_1}(\mathbf{t}_1 | \mathbf{x}_2) = e^{i\mathbf{t}_1^\top \boldsymbol{\mu}_1(\mathbf{x}_2)} \phi_{q(\mathbf{x}_2)}(\mathbf{t}_1^\top \Sigma_{11.2} \mathbf{t}_1).
\end{aligned}$$

A prova de i) e ii) pode ser encontrada em Arellano-Valle (1994) e Kelker (1970), respectivamente. Observe que $\alpha_{q(\mathbf{x}_2)} = -2\phi'_{q(\mathbf{x}_2)}(0)$ e no caso normal $\phi_{q(\mathbf{x}_2)}(u) = e^{-\frac{1}{2}u}$, $u \geq 0$, e conseqüentemente $\alpha_{q(\mathbf{x}_2)} = 1$ para todo $\mathbf{x}_2$.

## A.3 Distribuição $t$-Student

Nesta seção apresentamos um resumo das principais propriedades da distribuição $t$-Student utilizadas na demonstração dos teoremas que envolvem essa distribuição. Os resultados que apresentamos a seguir podem ser encontrados em Arellano-Valle e Bolfarine (1995).

**Definição A.3** *Um vetor aleatório* $\mathbf{X} : n \times 1$ *é dito ter distribuição t-Student n-variada com vetor de locação* $\boldsymbol{\mu} : n \times 1$*, matriz de dispersão* $\boldsymbol{\Sigma} : n \times n$ *não negativa e* $\nu$ *graus de liberdade, se*

$$\mathbf{X} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + V^{\frac{1}{2}}\mathbf{Z}, \tag{A.19}$$

*onde* $\mathbf{Z} \sim N_n(\mathbf{0}, \boldsymbol{\Sigma})$*,* $\mathrm{posto}(\boldsymbol{\Sigma}) = k$ *e* $V \sim \nu/\chi^2_\nu$ *são independentes e escrevemos por*

$$\mathbf{X} \sim t_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu) \tag{A.20}$$

Se $\mathrm{posto}(\boldsymbol{\Sigma}) = k = n$, isto é, $\boldsymbol{\Sigma} > 0$, então $\mathbf{X}$ tem densidade dada por

$$P_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = k(n, \nu)\nu^{\nu/2}|\boldsymbol{\Sigma}|^{-1/2}[\nu + (\mathbf{x} - \boldsymbol{\mu})^{\top}\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \mathbf{u})]^{-(\nu+n)/2}, \tag{A.21}$$

onde

$$k(n, \nu) = \frac{\Gamma(\frac{1}{2}(\nu + n))}{\pi^{\frac{1}{2}n}\Gamma(\frac{1}{2}\nu)}.$$

Em particular, se $\boldsymbol{\Sigma} = \mathbf{I}_n$, temos que $\mathbf{X}$ tem densidade dada por $f(\mathbf{x}^{\top}\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^n$, onde

$$f(u) = k(n, \nu)\nu^{\nu/2}(\nu + u)^{-(\nu+n)/2} \tag{A.22}$$

com $k(n, \nu)$ como em (A.21).

Se $\mathrm{posto}(\boldsymbol{\Sigma}) = k < n$, então a distribuição de $(\mathbf{X}|V = \nu)$ é $N_n(\mathbf{0}, \nu\boldsymbol{\Sigma})$ singular. Conseqüentemente, a distribuição condicional de $\mathbf{X}$ é também singular. Este caso pode ser interpretado como uma versão singular da distribuição $t$-Student e em tal caso é tratado similarmente à distribuição elíptica no caso singular (veja Cambanis et al., 1981).

Uma extensão da distribuição $t_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu)$, com $\boldsymbol{\Sigma} > 0$, pode ser obtida assumindo que em (A.19) a variável aleatória $V$ tem distribuição $\lambda/\chi_\nu \equiv GI(\lambda/2, \nu/2)$. Neste caso, $\mathbf{X}$ tem densidade dada por

$$P_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \lambda, \nu) = k(n, \nu)\lambda^{\nu/2}\{\lambda + (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\}^{-(\nu+n)/2}, \tag{A.23}$$

$\mathbf{x} \in \mathbb{R}^n$, onde $k(n, \nu)$ é como em (A.21). Nesta situação, escrevemos $\mathbf{X} \sim t_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \lambda, \nu)$ e diremos que $\mathbf{X}$ tem distribuição $t$-Student generalizada $n$-variada. Em particular, quando $\lambda = \nu$, temos a distribuição $t$-Student usual. Claramente, se $\mathbf{X}$ tem densidade dada por (A.23), então $\mathbf{X} \sim E\ell_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Assim, podemos pensar também em uma versão singular da distribuição $t$-Student generalizada, definida similarmente como em (A.19), com $V \sim \lambda/\chi_\nu \equiv GI(\lambda/2, \nu/2)$ e $\mathbf{Z} \sim N_n(\mathbf{0}, \boldsymbol{\Sigma})$, posto$(\boldsymbol{\Sigma}) = k < n$, independentes.

Seja $\mathbf{X} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + V^{1/2}\mathbf{Z} \sim t_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \lambda, \nu)$, com $V \sim \lambda/\chi_\nu$ e $\mathbf{Z} \sim N_n(\mathbf{0}, \boldsymbol{\Sigma})$ independentes. Então valem as seguintes propriedades:

i) $E[\mathbf{X}] = \boldsymbol{\mu}$, $\nu > 1$, $\mathrm{Cov}[\mathbf{X}] = \frac{\lambda}{\nu-2}\boldsymbol{\Sigma}$, $\nu > 2$;

ii) se $\boldsymbol{\eta} \in \mathbb{R}^n$ e $\mathbf{A}$ é uma matriz $m \times n$ (não estocástica), então

$$\boldsymbol{\eta} + \mathbf{A}\mathbf{X} \sim t_n(\boldsymbol{\eta} + \mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top; \lambda, \nu).$$

Em particular, se posto$(\boldsymbol{\Sigma}) = k = n$,

$$\boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu}) \sim t_n(\mathbf{0}, \mathbf{I}_n; \lambda, \nu).$$

iii) Partindo $\mathbf{X}$, $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ como

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \quad \text{e} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

onde $\mathbf{X}_1$ e $\boldsymbol{\mu}_2$ são vetores $n \times 1$ e $\boldsymbol{\Sigma}_{11}$ é uma matriz $n \times n$, então

$$(\mathbf{X}_1|\mathbf{X}_2 = \mathbf{x}_2) \sim t_n(\boldsymbol{\mu}_1(\mathbf{x}_2), \boldsymbol{\Sigma}_{11.2}; \lambda_{q(\mathbf{x}_2)}, \nu_m),$$

onde $\mu_1(x_2)$, $\Sigma_{11.2}$ e $q(x_2)$ são como no Teorema A.4 e os parâmetros $\lambda_a$ e $\nu_m$ são dados por

$$\lambda_a = \lambda + u, \; u \geq 0,$$

$$\nu_m = \nu + n - m, \; 1 \leq m \leq n.$$

Além disso,

$$\begin{aligned}
E[\mathbf{X}_1|\mathbf{X}_2] &= \mu_1(\mathbf{X}_2), \quad \nu_m > 1, \\
\mathrm{Cov}[\mathbf{X}_1|\mathbf{X}_2] &= \frac{\lambda_{q(\mathbf{X}_2)}}{\nu_{m-2}}\mathbf{\Sigma}_{11.2} \\
&= \frac{\lambda + (\mathbf{X}_2 - \mu_2)^{\top}\mathbf{\Sigma}_{22}^{-1}(\mathbf{X}_2 - \mu_2)}{\nu + n - m - 2}\left(\mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\right), \quad \nu_m > 2.
\end{aligned}$$

A afirmação seguinte será de grande utilidade na demonstração dos principais resultados no modelo funcional, quando os erros têm distribuição $t$-Student. Para sua demonstração, veja Arellano-Valle (1994).

**Lema A.1** *Seja* $\mathbf{X} \sim t_n(\mathbf{0}, \mathbf{I}_n; \lambda, \nu)$. *Então*

$$E[\|\mathbf{X}\|^{2k}(\lambda + \|\mathbf{X}\|^2)^{-m}] = \frac{B\left[\frac{1}{2}(p+2k), \frac{1}{2}(v+2m-k)\right]}{\lambda^{(2m-2k)/2}B\left(\frac{1}{2}p, \frac{1}{2}\nu\right)},$$

$v + 2m - 2k > 0$, *onde* $B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$ *é a função beta.*

# Bibliografia


ANDERSON, T.W.; FANG, K.T.; HSU, H. (1986). Maximum-likelihood estimates and likelihood-ratio criteria for multivariate elliptical conroured distributions. *The Canadian Journal of Statistics*, **44**, 55–59.

ANDERSON, T.W.; FANG, K.T. (1987). Cochran's theorem for elliptically contoured distributions. *Sankhyā* A, **49**, 305–315.

ABRAMOWITZ, M.; STEGUN, I.A. (1972). *Handbook of mathematical functions.* New York: Dover Publications, Inc.

ARELLANO-VALLE, R.B. (1994). *Distribuições elípticas: propriedades, inferência e aplicações a modelos de regressão.* Tese, IME-USP.

ARELLANO-VALLE, R.B.; BOLFARINE, H. (1995). On some characterizations of the $t$-distribution. *Statistics and Probability Letters*, **25**, 79–85.

ARELLANO-VALLE, R.B.; BOLFARINE, H. (1996). A note on the simple structural regression model. *Ann. Inst. Statist. Math.* **48(1)**, 111–125.

ARNOLD, S.F. (1981). *The theory of linear models and multivariate analysis.* New York: Wiley.

BARNDORFF-NIELSEN, O.E.; COX, D.R. (1984). Bartlett adjustments to the likelihood ratio statistics and the distribution of the maximum likelihood estimator. *Journal of the Royal Statistical Society B*, **46**, 483–495.

BARNETT, V.D. (1969). Simultaneous pairwise linear structural relationships. *Biometrics*, **25**, 129–142.

BERKANE, M.E.; BENTLER, P.M. (1987). Characterizing parameters of multivariate elliptical distributions. *Communication Statistic - Simulation*, **16**, 193–198.

BOLFARINE, H.; CORDANI, L.K. (1993). Estimation of a structural regression model with known reliability ratio. *Annals of the Institute of Statistical Mathematics*, **45**, 3, 531–540.

BOLFARINE, H.; RODRIGUES, J.; CORDANI, L.K. (1992). Modelos com erros nas variáveis. *ABE*.

BOLFARINE, H.; NASCIMENTO, J.A. (1992). *Bartlett correction factors for the structural regression model with known reliability ratio.* São Paulo, IME-USP, 1992 (RT-MAE-9206).

BOLFARINE, H.; GALEA-ROJAS, M. (1995). Comments on "Functional Comparative Calibration using on EM Algorithms" by D. Kimura. *Biometrics*, **51, 4**, 1579–1580.

BROWNE, M.W. (1982). *Covariance structures.* In topics in Applied Multivariate Analysis, ed. D.M. Hawskins, Cambridge, U.K.: Cambridge University Press, 72-141.

BROWNE, M.W. (1984). Asymptotically distribution-free methods for analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, **37**, 62–83.

BUONACCORSI, P. (1989). Errors-in-variables with sistematic biases. *Commun. Statist. - Theory Meth.*, **18(3)**, 1001–1021.

CAMBANIS, S.; HUANG, S.; SIMONS, G. (1981). On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis*, **11**, 368–385.

CHENG, C.; VAN NESS, J. (1991). On the unreplicated ultraestrutural model. *Biometrika*, 442–445.

CORDEIRO, G.M. (1987). On the corrections to the likelihood ratio statistics. *Biometrika*, **74**, 265–274.

CORDEIRO, G.M.; FERRARI, S.L.P. (1991). A modified score statistic having chi-squared distribution to order $n^{-1}$. *Biometrika*, **78**, 573–582.

COX, D.R.; REID, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *Journal of the Royal Statistical Society B*, **49**, 1–39.

DEVLIN, S.J.; GNANADESIKAN, R.; KETTERING, J.R. (1976). Some multivariate applications of elliptical distributions. In: *Essays in Probability and Statistics* (Eds. S.I. Keda et al.), Chapter 24, Tokyo: Shinko Tsusho Co., 365–393.

DOLBY, G.R. (1976). A ultraestructural relation: a synthesis of the functional and structural relations. *Biometrika*, **63**, 39-50.

DUNN, G. (1992). *Design and analysis of reliability: the statistical evaluation of measurement errors.* Edward Arnold. New York.

FANG, K.T.; KOTZ, S.; NG, K.W. (1990). *Symmetrik multivariate and related distributions.* London-New York: Chapamn and Hall.

FANG, K.T.; ZHANG, Y.T. (1990). *Generalized Multivariate Analysis.* Springer-Verlag.

FLEMING, W. H. (1966). *Functions of several variables.* Addison-Wesley.

FULLER, W. (1987). *Measurement error models.* New York: Wiley.

GHOSH, M.; SINHA, K.B. (1980). On the robustness of least squares procedures in regression models. *Journal of Multivariate Analysis*, **10**, 332–342.

GLESER, L.J. (1981). Estimation in a multivariate error-in-variables regression model: large sample results. *Annals of Statistics*, **9**, 24–44.

GLESER, L.J. (1985). A note on G.R. Dolby's unreplicated ultraestructural error in variables model. *Biometrika*, **72**, 117–124.

GRAYBILL, F.A. (1983). *Matrices with applications in statistics.* 2.ed. Belmont, California: Wadsworth.

GRUBBS, F.E. (1948). On estimating precision of measuring instruments and product variability. *Journal of the American Statistical Association*, **43**, 243–264.

GRUBBS, F.E. (1973). Errors of measurement, precision, accuracy and the statistical comparison of measuring instruments. *Technometrics*, **15**, 53–66.

HAYAKAWA, T. (1977). The likelihood ratio criterion and the asymptotic expansion of its distribution. *Ann. Inst. Math.*, **29**, Part A, 359–378.

HENDERSON, H.V.; SEARLE, S.R. (1979). Vec and Vech operators for matrices, with some uses in Jacobians and multivariate statistics. *The Canadian Journal of Statistics*, **7**, 65–81.

KAAK, R.; RIBOLI, E.; ESTIVE, J., VAN KAPPEL; VAN STAVEREN (1994). Estimating the accuracy of questionnary assessments: validation in terms of structural equation models. *Statistics in Medicine*, **13**, 127–142.

KANO, Y.; BERKANE, M.; BENTLER, P. (1993). Statistical inference based on pseudo-maximum likelihood estimators in elliptical populations. *Journal of the American Statistical Association*, **88, 421**, 135–143.

KARIYA, T.; EATON, M.L. (1977). Robust tests for spherical symmetry. *The Annals of Statistics*, **5(1)**, 206–215.

KELKER, D. (1970). Distribution theory of special distributions and a location-scale parameter. *Sankhyā A*, **32**, 419–430.

KELLY, G.E. (1984). The influence function in the errors in variables problem. *The Annals of Statistics*, **12**, 87–100.

KELLY, G.E. (1985). Use of the structural equations model in assessing the reliability of a new measurement technique. *Applied Statistics*, **34**, 258–263.

KELLY, G.E. (1992). Robust regression estimators – the choice of tuning constants. *The Statistician*, **41**, 303–314.

KENDALL, M.G.; STUART, A. (1979). *The advanced theory of statistics.* 4.ed. New York: Hafner.

KIEFER, J.; WOLFOWTZ, A. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.*, **27**, 887–906.

KIMURA, D. (1992). Functional comparative calibration using the EM-algorithm. *Biometrics*, **48**, 1263–1271.

LANGE, K.L.; LITTLE, J.A.; TAYLOR, M.G. (1989). Robust statistical modeling using the $t$-distribution. *Journal of the American Statistical Association*, **84(408)**, 881–896.

LAWLEY, D.N. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika*, **43**, 295–303.

LINDER; BABU, J. (1994). Bootstrapping the linear functional relationship with known error variances ratio. *Scandinavian Journal of Statistics*, **21(1)**, 21–40.

LITTLE, R.J.A. (1988). Robust estimation of the mean and covariance matrix from data with missing values. *Appl. Statist.* **37(1)**, 23–38.

MAK, T. (1982). Estimation in the presence of incidental parameters. *Canadian Journal of Statistics*, **10**, 121–132.

McCULLAGH, P.; COX, D.R. (1986). Invariants and likelihood ratio statistics. *The Annals of Statistics*, **14**, 1419–1480.

MITCHEL, A.F.S. (1988). Statistical manifolds of univariate elliptic distributions. *International Statistical Review*, **56(1)**, 1–16.

MITCHEL, A.F.S. (1989). The information matrix, skewness tensor and $\alpha$-connections for the general multivariate elliptic distribution. *Ann. Inst. Statist. Math.*, **41(2)**, 289–304.

MUIRHEAD, R.J. (1980). The effects of elliptical distributions on some standard procedures involving correlation coefficients: a review. *Multivariate Statistical Analysis*, 143–159.

MUIRHEAD, R.J. (1982). *Aspects of multivariate statistical theory.* New York: Wiley.

MUIRHEAD, R.J.; WATERNAUX, C.M. (1980). Asymptotic distributions in canonical correlation analysis and other multivariate procedures for nonormal populations. *Biometrika*, **67(1)**, 31–43.

NASCIMENTO, J. (1994). *Parametrização ortogonal e outras contribuições ao modelo com erros nas variáveis.* Tese, IME-USP.

NEYMAN, J.; SCOTT, E.L. (1948). Consistent estimates based on partially consistent obsrvations. *Econometric*, **16**, 1–16.

NEYMAN, J.; SCOTT, E.L. (1951). On certain methods of estimating the linear structural relation. *Ann. Math. Statist.*, **22**, 252–261. Correction, *Ann. Math. Statist.* (1952), **23**, 115.

PATEFIELD (1976). On the information matrix in the linear functional relationship problem. *Applied Statistics*, **26**, 69–70.

SEN, P.K.; SINGER, J.M. (1993). *Large sample methods in statistics: an introduction with applications.* Chapman & Hall. New York.

SHAPIRO, A.; BROWNE, M.W. (1987). Analysis of covariance structures under elliptical distributions. *Journal of the American Statistical Association*, **82(400)**, 1092–1097.

SPRENT, P. (1966). A generalized least-squares approach to linear functional relationships. *Journal of the Royal Statistical Society, B*, **28**, 278–297.

TAYLOR, J.M.G (1992). Properties of modelling the error distribution with an extra shape parameter. *Computational Statistics & Data Analysis*, **13**, 33–46.

TYLER, D.E. (1982). Radial estimates and the test for sphericity. *Biometrika*, **69(2)**, 429–436.

WILLIAMS, E.J. (1969). Regression methods in calibration problems. *Bulletin of the International Statistical Institute*, **43**, 17–28.

WONG, M.Y. (1989). Likelihood estimation of a simple linear regression model when both variables have error. *Biometrika*, **76(1)**, 141–148.

WONG, M.Y. (1991). Bartlett adjustment to the likelihood ratio statistic for testing several slops. *Biometrika*, **78**, 121–124.

YAMAGUCHI, K. (1990). Generalized EM algorithm for model with contamined error term. In *Proceedings of The Seven Japan and Korea Joint Conference of Statistics*, 107–114.

ZELLNER, A. (1971). *An introduction to Bayesian inference in econometrics*. New York: Wiley.

ZELLNER, A. (1976). Bayesian and non-Bayesian analysis of the regression model with multivariate Student-$t$ error terms. *Journal of the American Statistical Association*, **71(354)**, 400–405.